DYOLOGY

On Diversity and Equality

VOLUME II

Ethics and Political Philosophy

ROBIN MAINS



D Y O L O G Y

Dyology: On Diversity and Equality

Volume I · Metaphysics and Philosophy of Religion

Volume II · Ethics and Political Philosophy

Volume I contains the General Introduction, Chapters 1–12 and Appendices A–C; Volume II contains Chapters 13–24 and the General Conclusion. Chapter numbers run through both volumes. The Premise Register and the Glossary are printed in each, and each volume's Bibliography lists the works cited in it.

D Y O L O G Y

On Diversity and Equality

VOLUME II

Ethics and Political Philosophy

ROBIN MAINS



Dyology: On Diversity and Equality
Volume II: Ethics and Political Philosophy

© 2026 Robin Mains. All rights reserved.
Pre-publication manuscript, 2026.

ISBN (two-volume set): to be assigned
ISBN (Volume I): to be assigned
ISBN (Volume II): to be assigned

This book was written with the assistance of large language models; the Note to Volume II says how.

For correspondence: robinmains.author@gmail.com

Contents

NOTE TO VOLUME II.....IX
GENERAL INTRODUCTION.....1
13. CONSCIOUSNESS AS DIFFERENTIATION.....19
14. EMOTION AS VALUE.....59
15. EVALUATION AND CHOICE.....139
16. THE GREAT MORAL CONFLATION.....167
17. CERTAINTY, PERFECTION, AND DOGMA.....205
18. FREE WILL AND THE PERSPECTIVAL ASYMMETRY.....235
19. THE CONSTITUTION OF PURPOSE.....279
20. PURPOSE AND PUNISHMENT.....309
21. CONFLICT AND DEEP PERSUASION.....343
22. DEMOCRACY AND THE LAWMAKING MECHANISM.....375
23. THE DEMOCRATIC DEFICIT OF THEOCRACY.....423
24. PEACE AND GOD.....457
GENERAL CONCLUSION.....495
PREMISE REGISTER.....513
GLOSSARY.....541
BIBLIOGRAPHY.....577

VOLUME I

AUTHOR'S NOTE
GENERAL INTRODUCTION
1. THE SELF-VINDICATION OF DIFFERENCE
2. THE CONTRASTIVE STRUCTURE OF CONTENT
3. THE EXHAUSTIVE PARTITION
4. THE UNDIFFERENTIATED LIMIT OF GROUND
5. THE UNDIFFERENTIATED BOUNDARY OF TIME
6. COGNITION AND ZERO COMPLEXITY
7. THE GRID OF MEANING
8. THE ORIGINAL ILLUSION
9. THE DISSOLUTION OF CREATION
10. THE SEPARATION OF GROUND AND MIND
11. EQUALITY AND THE KALĀM'S GOD
12. THE LIMITS OF NEGATIVE THEOLOGY
APPENDIX A: WHY THERE IS SOMETHING RATHER THAN NOTHING
APPENDIX B: THE TERMINUS WITHOUT DIVERSITY
APPENDIX C: THE GRIDS OF ANALYTIC PHILOSOPHY
PREMISE REGISTER
GLOSSARY
BIBLIOGRAPHY

NOTE TO VOLUME II

Should you read this?

You are about to decide whether this book is worth your time, and I think you should be able to work that out in ten minutes rather than in a thousand pages. This note is here to make that decision easier for you.

It is also here because this book was built with heavy use of AI, and I would much rather tell you that myself. Among other debated issues, which I will come to later, one of the greatest impoverishments of AI-generated text is the absence of the warmth of a human mind, which inevitably comes through the pages of even a dry academic text.

This is the second of two volumes. The first laid the bedrock, that at least one difference obtains, and carried it through the metaphysics of difference to the ground and to the God that was supposed to stand there; this one carries it through value and morality to peace. The Author's Note that opens Volume I tells the story of the book and of the unusual way it was built. What you need to know from it before reading this volume is repeated here.

Why Dyology?

It was only an accident that the publication of *Zero Complexity* coincided with the explosive advent of AI. I could not resist testing my lifelong project, and of course the first thing I asked was, "what do you think about the philosophy of this book?" That was the beginning of a second journey, which led to Dyology, a book I often describe to my friends and family as "Zero Complexity on steroids".

This is to say that the core framework itself is mine, resting on premises and derivations that I put together, from first metaphysical principles all the way to political considerations, well before I engaged with the academic literature and before any AI existed that could have built it for me. The 2025 book is the dated public record of that, and you are welcome to check it. It was itself written without AI, which entered it only for copyediting, for a quarantined preliminary appendix on the literature, and for an epilogue reporting a dialogue with an AI about AI.

Starting from *Zero Complexity*'s framework, I provided the complete source book to Claude, a large language model developed by Anthropic, with the work spanning Claude Opus 4.6 through Opus 5, and Claude Fable 5.1. I set it three tasks: to extrapolate the main claims and position them within the relevant academic domains; to subject those claims to the most candid and hostile engagement with every relevant philosophical paper or tradition the machine could think of; and to check the consistency of the architecture across all twenty-four chapters. Each chapter is indeed built that way: a claim, several objections, and a reply to every one of them.

Four hours, and job done? Not quite. This has taken thousands of hours; it has been a brutal, iterative process. I was pleased to see that the armchair framework I built over two decades was very resilient, yet there were several vulnerabilities that AI highlighted, and I had to deal with them. But how could I possibly respond to Derrida, Hegel, Schaffer, Wittgenstein, Kant? I asked to be treated like a ten-year-old: "can you explain to me, in very simple terms, what the vulnerability is, and what I need to figure out in order to solve it?" Sometimes the discussion would lead to a stalemate; sometimes to a victory. Even the most specialised philosophical position can be explained in plain terms, as in medicine, and once I had the core of an objection I could usually reply: either by confirming a position already implicit in what I had developed, or by building something new. The clearest and most prolonged case is Chapter 14, on metaethics, where I was put up against the most hostile realist objections. Two examples from my side of that exchange will give you the grain of it: the vacancy reading of the robust realist's clause (Chapter 14, §§3.6–3.7, A.1–A.4), and the anchoring of moral value in the affected party's own registration, which owes something to the bedside — the fact of a patient's suffering sits in the patient, whatever the room believes about it.

So, if this book were a creature, an animal, it would have the skeleton of *Zero Complexity* and the muscle and flesh of that long argument.

Are the references real, or hallucinated?

The obvious danger, and one you are right to worry about, is that the animal is a piece of taxidermy — beautifully AI-wrapped but empty of substance — or, worse, a piece of holographic taxidermy: philosophy that is not only incoherent and unoriginal but built on engagement with counterparts that are inaccurately represented, or outright fabricated. But I doubt it. There are 736 references across the two volumes of this book, and I have spent many hours and many tokens verifying them with successive models, most repeatedly and with real web searches. I checked them against the characteristic failure modes of AI-assisted scholarship: citations that do not exist, or that do not say what they are cited as saying; positions mischaracterised; objections answered in a weakened form. The exercise also revealed that several pieces of published pre-AI literature contain inaccuracies of their own, which made it harder than I expected. If you find one of those failure modes here, you have found a real defect, and the Premise Register will show you what falls with it.

This book is built to be attacked

My greatest preoccupation in writing *Zero Complexity* was to be understood, to be clear, but most importantly to own my claims. With Dyology I kept the same values, but given the enormous amount of philosophy that AI has engaged with, I simply could not own every secondary claim and derivation. I could assess the faithfulness of any application offered to me, but not in that much detail. So I asked AI to audit and register the strength of every claim, so that each is auditable.

Put simply: the metaphysics of the earlier chapters supports the metaethics of the later ones, and that dependency has been tracked — obsessively, I would say — across the entire book. If you don’t buy the claim in Chapter 2, you do not need to bother with Chapter 4, as it depends on it. Conversely, you may start with Chapter 12, but if you wanted to explore the metaphysical foundation of its claims, you would have to go back to the earlier chapters.

In fact, AI has managed to make this framework as vulnerable to criticism as possible, which is what I wanted. There is no point in looking right as a function of obscurity. I would much rather be wrong and learn the truth from someone else, and clarity is the only way I can make sure that someone is able to correct me. There is also a great deal the machine cannot do; it matters more than what it can. It cannot establish that a premise is true, that a result is significant, or that the objection it has built is the strongest a specialist would put to me, and its simplifications can soften an objection in ways neither of us would notice. Your scrutiny is the tribunal it cannot replace. My gamble, of course, is that I am right, until proven otherwise.

If you are still here, where to start?

Triage me. Upload a chapter to whichever model you prefer, and ask it to find the weakest link; it will cost you ten minutes and very little money. Some might dislike, and even condemn, the use of AI for triaging a text, but I find it hard to believe that it is not already widely used for this very purpose, whether publicly declared or not. Its verdicts need to be studied before they can be read properly. I would encourage you to get to know this new technology, even if it is only a matter of "keeping your enemies closer".

If you are a specialist, I would recommend reading the General Introduction, finding on its map of the argument (§3 and Figure 1) which part of the framework deals with your field of expertise, and triaging the relevant chapter. If it doesn't sound like crackpot philosophy, give it a go. Maybe check a handful of references while you are there.

Why this is not a finished book

This book needs credibility; that is something I cannot provide. Only real philosophers can gift this framework with that. Does the truth need authority and credentials? No. The truth is the truth, whether it comes from a child, a doctor, a philosopher or a machine, whether it arrives accidentally or through rigorous reasoning. But the reassurance

that a text is not a well-written waste of time: well, yes, that does require credibility.

That is why I am not putting Dyology out as a finished thing. My hope is that a professional philosopher who has not yet learnt to hate AI indiscriminately, and who has had the benevolence to read this far, will also have the curiosity to test it, read a chapter or two, and conclude that this manuscript might be a valuable contribution, and might deserve to be read by others.

And if you find yourself working on a chapter with me — extending it, correcting it, taking it somewhere I could not have taken it on my own — then that is real collaboration, and the chapter should carry your name next to mine on the strength of that work. I would remain the author of the book; the chapter would be ours. I do not believe there are many projects of this kind out there. It would closely resemble, in nature, many of the medical textbooks I am familiar with, where specialists from different fields build something new together.

Finally, let's put ourselves in each other's shoes. Yes, AI is the most disruptive and corruptive piece of technology we have seen. Yes, PhilPapers has been inundated by "independent" researchers. I give you all that. What the flood has done is put shiny glass and cut stone in the same post. Telling one from the other takes training I do not have and ten minutes of someone who does. I am asking for the ten minutes, knowing that the odds are on the glass.

If you find any value in what you read next, and wish to contribute to this book, email me at robinmains.author@gmail.com.

With thanks for your time,

Robin Mains
London, 2026

General Introduction

1 From Difference to Peace

Sir Peter Mansfield and Paul Lauterbur shared the Nobel Prize in Physiology or Medicine for the work that made Magnetic Resonance Imaging possible: a tool now used millions of times a year to diagnose disease. Anyone can watch the image of a brain appear on a screen, while only a handful of scientists could explain every physical principle that puts it there, down to the most intricate quantum calculation.

Philosophical research works in much the same way. Once a good idea comes to light, anyone can hold it and use it, but not everyone will know its roots — why it is sound, and why it stays alive inside a web of objections, doubts, and the ever-present risk of fallacy.

This book puts forward a set of ideas about how to resolve conflict: in the broad sense of collective peace, and in the closer sense of understanding ourselves, our friends, and our families. Conflict is a source of suffering second only to disease, and as a doctor I felt compelled to investigate a form of pain that no surgery or analgesia can reach.

Philosophy is, in a way, the physics of politics, and politics is what carries a society towards peace or away from it. Just as Mansfield's equations let us see the minute detail of a living brain, a sound set of ideas might let us see our way past war and humanitarian catastrophe.

This is the question the book sets out to answer. Is it possible to build a rational bedrock for the resolution of conflict — something formidably resilient and never rigid, the opposite of what absolutist and relativistic doctrines offer? The answer has to begin somewhere that will hold, since everything after it depends on how well that first stone is laid.

Every philosophy has a first move, and most can be refused. A reader who does not care for the premise simply declines it, and everything built on it comes down with a shrug — *it sounds reasonable, but it could be different; you never know*. That small remark is fatal; it is fatal because it costs nothing to make.

A framework answerable to it can be argued with, but it cannot be relied on. It floats.

So the first thing this book needed was a premise that could not be refused in that way: not merely a plausible premise, or a widely shared one, or one that is hard to imagine denying, but one whose denial defeats itself in the act. A premise of that kind is a fixed point. Whatever else is disputed, it stays where it is, and a reader who wants to reject what follows must find the error further down the chain, in a specific inference, and say which one. That is what makes the rest of the book worth arguing with instead of merely worth doubting.

The premise that cannot be refused. There is such a premise; it is unglamorous: *at least one difference obtains*. Try to deny it, sincerely, and the denial you have just formed differs from its own negation, from silence, from every other thing you might have said instead. The denial exhibits what it denies. It is a floor, and a thin one: one difference and no more, a result Chapter 1 accepts as shallow in the deflationist's sense. Its worth is its security. Chapter 1 states it rigorously and defends it against the objections it has to survive; for now, what matters is that the foundation does not have to be granted, because it cannot be coherently refused.

What follows is a long walk from that fact, and it goes further than a reader might expect: through meaning and mind, through the terminus of explanation and the God that was supposed to stand there, through value and morality, to the conditions under which a conflict can be resolved instead of managed. Every step is graded for how much weight it will bear; the grades thin as the walk climbs, from an established footing to a proposed keystone. The walk has a shape, which §3 draws and grades part by part before the map is filled in.

2 A Note on the Title

The word *Dyology* is coined from the Greek *dyo* (two) and *logos* (reason or account). The name carries the framework's structure at both of its scales. At the foundation, the minimal two: a difference is one of two, this and what it excludes, and the book's single undeniable premise

— that at least one difference obtains — is the smallest duality there is. At the summit, the great two: diversity and equality, the differentiated and the undifferentiated, the pair that exhausts the space of metaphysical possibility. The two readings are one structure at two scales; deriving the second from the first is the work of the opening chapters.

3 The Anatomy of the Book

The twenty-four chapters have the shape of an arch (Figure 1). The bedrock is difference, the one thing that cannot be denied. On it sits a footing, ontological and then epistemological. From the footing two sides rise, one theological and one metaethical, each as a pier and then a curve — a core, and the application that turns the side toward the other — and each the removal of an authority that could otherwise settle a dispute from outside: the theological side establishes that no mind stands above the hierarchy of difference, the metaethical side that no moral order stands outside the parties. The keystone is peace, the last stone set and the one neither side can carry alone (Chapter 24): on this account peace comes only through deep persuasion, the transformation of the preferences that generated a conflict, since no authority outside a dispute remains to settle it; the God that survives the theological side is the one who persuades and does not command.

Each chapter exists to establish one central claim, a proposition about the world; the results it adds in establishing that claim, its supporting results, are graded on their own in the Premise Register, whose rows are grouped under the same claims (§6). The table below gives the twenty-four claims from bedrock to keystone, in the arch's order, with the position elsewhere in philosophy that each underwrites; the paragraphs after it read the arch in the same direction.

Most of the positions the claims underwrite are held already — hard incompatibilism, sentimentalism about value, the case against divine simplicity, the temporal God of open theism, the impossibility of nothing, the convergence of the mystical traditions, the fallibilist case against dogma — and that the framework lands on positions others

reached by other roads is the strongest evidence the book has for it. What each supporting result adds is of one of four kinds: a mechanism, which explains why something holds and is original as insight; an instrument, which reaches a conclusion by a new route and is original as proof; a distinction, which draws a new cut and is original as taxonomy; a diagnosis, which reads many positions through one lens and is original as synthesis. The Premise Register is a fifth kind, original as method (Appendix C, §7). Which kind each result is, and what would refute it, is stated where the result is made.

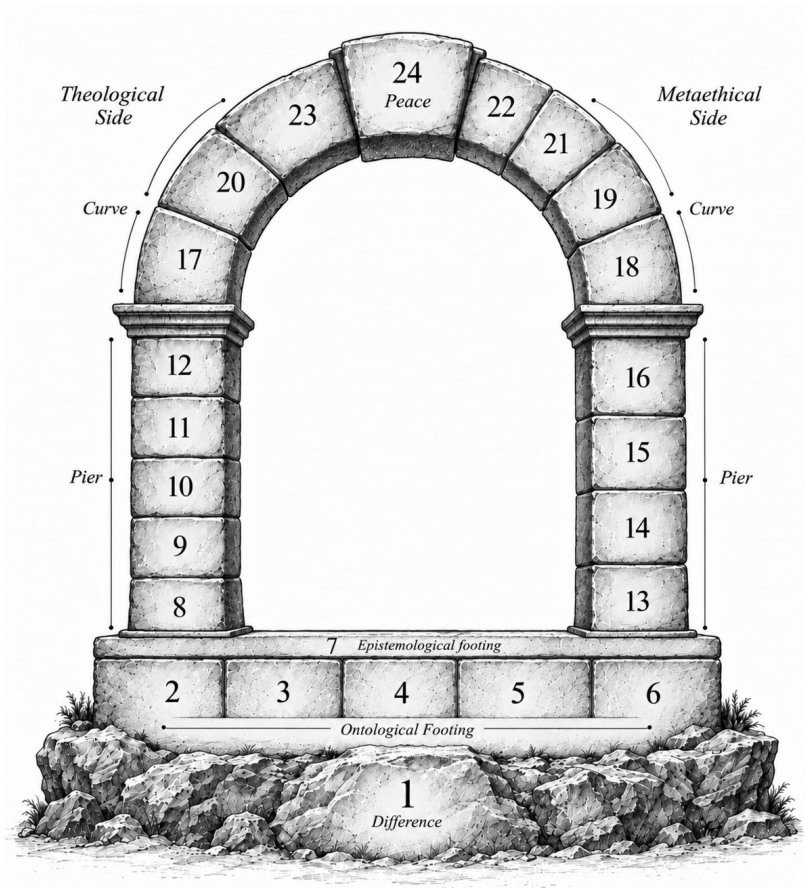


Figure 1. The arch of the argument: bedrock (1), footing (2–7), the theological side (pier 8–12, curve 17, 20, 23), the metaethical side (pier 13–16, curve 18, 19, 21, 22), the keystone (24). Load, not reading order.

Part of the arch	Claim (chapter)	Position it underwrites
Bedrock	At least one difference obtains: the one thing that cannot be denied. (1)	The reality of difference, against the sceptic (Stroud)
The footing (Chapters 2–7).		
Footing: ontological (2–6)	Content requires contrast, and the diff is the unit of content. (2)	Contrastive accounts of content (Saussure, Brandom, Dretske)
	The diversity/equality partition is exhaustive. (3)	Every account that ends in an undifferentiated ground
	Equality is the foreclosing terminus of grounding. (4)	Termination in grounding (foundationalism), on the constitutive order
	Time terminates at the same boundary, and the two hierarchies converge. (5)	Kant's antithesis; time with a boundary and no first moment
	The foreclosing terminus is unintelligent. (6)	The lower bound every structural theory of consciousness assumes
Footing: epistemo-logical (7)	Meaning and difference are one closed grid. (7)	The impossibility of ineffable content (the resolute Wittgenstein)

The theological side: no mind stands above the hierarchy of difference (Chapters 8–12, 17, 20, 23).		
Pier: core theology (8–12)	"Infinite" names two opposite conditions, and the terminus has three names. (8)	Negative theology's limit; the convergence of the mystical traditions (Stace, Forman)
	Creation is dependence, and the sustainer survives the transition picture. (9)	Creation as dependence (Aquinas; Grünbaum against the transition picture)
	Ground and mind come apart: the simple God is incoherent, the determinate God an inhabitant. (10)	The case against divine simplicity (Plantinga, Mullins)
	The kalām reaches no necessary being. (11)	The kalām cosmological argument (Craig)
	Negative theology names a limit of thought and misreads it as a being. (12)	Apophatic theology's recognition of a limit (Pseudo-Dionysius, Maimonides)
Curve: applied theology (17, 20, 23)	Perfection does not entail infallibility; dogma has a four-stage anatomy with a refutation condition. (17)	The fallibilist case against dogma (Mill, Küng)
	Desert fails by two routes, so restoration replaces retribution. (20)	Restorative and anti-retributive justice (Pereboom; communicative theories)

	Creed-based judgement is the conflation projected onto earthly standing: the divine wall. (23)	Reform theology's foundation/ application distinction (An-Na'im, Volf, Borowitz)
The metaethical side: no moral order stands outside the parties (Chapters 13–16, 18–19, 21–22).		
Pier: core metaethics (13–16)	The mental has two axes, content and orientation, and value lies on the second. (13)	Graded taxonomies of mind, from registration to reflection
	Value is constituted by the evaluative modality's response and anchored in the affected party; the moral is the empathic evaluation of chosen behaviour. (14)	Sentimentalism and response-dependence (Hume, Prinz, D'Arms and Jacobson); moral error without realism (Blackburn)
	Total value composes primary and added value; across evaluators a trade-off has no fact beyond convergence under mutual registration. (15)	The association evidence on preference (Bargh)
	The major ethical traditions mistook conformity to a standard for empathic evaluation. (16)	Sentimentalism's charge against rationalism (Hume to Prinz)
Curve: applied ethics (18, 19, 21, 22)	Authorship is bounded: freedom is real from the first-person present and determined from the third-person retrospect. (18)	Hard incompatibilism (Pereboom)

	A purpose is an object with a sense; it requires no endorsement from outside the field. (19)	Hybrid accounts of meaning in life (Wolf); atelic activity (Setiya)
	Structural peace comes only through deep persuasion, the transformation of the preferences that generated the conflict. (21)	The contact hypothesis and deep canvassing (Allport; Broockman and Kalla)
	The conditions for deep persuasion are analytic and its institutions means; relativism and absolutism share the authority-dependence that disables lawmaking. (22)	Deliberative democracy and free speech (Mill)
Keystone: peace (24)	The God that survives persuades and does not command: the companion. (24)	Process theology and open theism (Hartshorne, Hasker)
	Peace, where the two sides meet. (21–22, 24)	The aim; reached from the metaethical side in Chapters 21–22 and locked in Chapter 24

The bedrock. At least one difference obtains, since the denial of difference instantiates what it denies, a fourth mechanism of self-refutation, content-individuation (Chapter 1). This is the bedrock, the one claim that cannot coherently be denied, on which everything above it rests, at the grades §5 assigns.

The footing. Content requires contrast: to be one way is to be "one way rather than another," so a content with no contrast is no content, and the diff is the unit (Chapter 2). The diversity/equality partition is exhaustive, every state of reality being either differentiated or undifferentiated (Chapter 3). If the hierarchies of grounding and time

terminate, they do so in equality, a state of zero complexity at the structural boundary of both, which no cognitive property can characterise (Chapters 4–6). The grid of meaning is closed, since any counterexample to the biconditional of meaning and difference arrives already gridded (Chapter 7). The question "why is there something rather than nothing?" names by "nothing" a condition that cannot obtain, so what remains is why being is differentiated, answered as a brute fact about obtaining and owned where it falls (Chapters 8 and 9).

The theological side. Its pier, the core (Chapters 8–12): "infinite" names two opposite conditions, and the terminus has three names. Creation pictured as a transition from non-being is incoherent, while the mature doctrine of dependence survives under the name the sustainer. Ground and mind come apart: the God of classical theism read as absolutely simple conjoins what the structure of difference makes incompatible, and read as maximally determinate stands as an inhabitant of the diversity it was meant to ground. The kalām reaches no necessary being. Negative theology names a limit of thought and misreads it as a being. Its curve, the application that turns the side toward the other (Chapters 17, 20, 23): perfection does not entail infallibility, so no text or authority stands beyond revision; purpose requires neither God nor immortality; desert fails, so restoration replaces retribution; and creed-based judgement of persons is the conflation projected onto earthly standing, the divine wall. Together with the unintelligence of the terminus, these establish that no mind stands above the hierarchy of difference — the removal of the authority a creed appeals to.

The metaethical side. Its pier, the core (Chapters 13–16): the mental has two axes, content and orientation, and value lies on the second. Value is constituted by the evaluative modality's response to non-evaluative input, as colour is by the visual system's response to wavelength, and the moral is the empathic evaluation of chosen behaviour, its engagement with the emotional consequences of what is done for others; the moral fact is anchored in the affected party's own registration, a fact observer-independent though not mind-independent, so a verdict can be wrong without being fixed by anything beyond the parties. Total value composes primary and added value. The major

ethical traditions mistook conformity to a standard for this engagement. Its curve, the application that turns the side toward the other (Chapters 18, 19, 21, 22): authorship is bounded. A purpose is an object with a sense. Structural peace comes only through deep persuasion, and its conditions follow from what deep persuasion is. Together these establish that no moral order stands outside the parties — the removal of the authority the absolutist appeals to.

The keystone (Chapter 24). The two sides meet in peace. From the metaethical side, deep persuasion, the transformation of the evaluative preferences that generated a conflict, is the only route to structural peace because no authority outside a dispute remains to settle it; the conditions for it — expressibility, receptivity, and non-exclusion — are analytic and conditional on the aim, and free speech, open speech, and democratic equality are the best available means of securing them (Chapters 21–22). From the theological side, the God that survives the critique is a companion within existence who persuades and does not command, the book's most explicit proposal, with its posits costed where they are made. The keystone is where the two arrive together: the companion is the theological counterpart of deep persuasion, so the stone neither side can carry alone is the one both reach.

The arch thins as it rises. The bedrock is self-vindicating and the footing established; the theological side is established in its pier and, in its curve, established where it classifies and constructed where it applies; the metaethical side is constructed throughout, analytic where its political crown claims necessity and conditional on the aim; the keystone is established in what it eliminates and proposed in the God it keeps. That thinning is where the General Conclusion looks for the faults (§2), where the load-bearing joints are named.

Three appendices close Volume I. **Appendix A** collects, in one place and in plain order, the book's answer to the oldest question in metaphysics — why there is something rather than nothing — which the chapters give one step at a time. **Appendix B** takes the terminus without diversity through the three names the traditions gave it — one, zero, infinity — and finds one condition beneath them. **Appendix C** turns the

framework on analytic philosophy itself, reading its disputes as damage to the shared structure of meaning, and its schools as the repairs attempted on that damage, and closes by presenting the Premise Register as the method the book has used throughout.

Three further pieces of apparatus support the argument without forming part of it. The **Premise Register**, after the General Conclusion, collects every result the book establishes into a single table, with its grade and the results it rests on; each chapter's own Grades and Dependencies section justifies the grades locally. The **Glossary**, at the back, defines every recurring term in full. The **Bibliography** of each volume gathers the works cited in it, 736 across the two, which each chapter also lists separately at its own end.

4 The Anatomy of a Chapter

Every chapter is built to the same plan, so that a reader can find the load-bearing parts of an unfamiliar one in a few seconds.

The table gives the plan: each part, where it sits, and what it does.

Part	Where	What it does
Dependency statement	Under the title	Two lines: what the chapter presupposes, and what it establishes
In brief	Boxed, beneath the dependency statement	The argument in a paragraph, for a reader deciding whether to enter
Introduction	§1	Sets out the problem and the chapter's claims
Numbered sections	§2 onward	Develop the claims

Part	Where	What it does
Interim summary	Before the objections	"Where the argument stands": the results so far in one paragraph, so that the objections are met against a stated position
Objections and replies	Penultimate section	The strongest challenges the literature affords, each with an answer
Conclusion	Final section	States what has been established and at what cost
Grades and Dependencies	After the conclusion	Justifies each grade, records what a result inherits and where its exposure splits, notes where an instrument is later used, and states what a challenger would have to produce
Appendix	Chapters 1, 3, 4, 8, 13, 14, 16, 22 and 24 only	A formal statement, a model, or a survey of a literature that would have interrupted the argument
References	Last	The chapter's own works cited

5 A Grading System for Each Claim

The Grades and Dependencies section that closes each chapter assigns a grade to every result the chapter establishes, drawn from a four-term ladder; the Premise Register collects the grades. A **self-vindicating** result cannot be coherently denied, since the denial instantiates it. An **established** result is validly derived from defended premises. A **constructed** result is the best available positive account, with live rivals still standing. A **proposed** result is offered as explicitly revisable.

The variation along that ladder is deliberate: a framework claiming equal certainty for all its results would be dishonest. Two premises are

self-vindicating: the occurrence premise of Chapter 1, that an entertaining is an occurrence, and the composition premise beside it, that an entertaining of the denial has at least two components. So is the foundational result built from the two alone, that at least one difference obtains. The vindication law — the necessary conditional through which Chapter 1's mechanism runs — and the exhibitive closure of meaning (Chapter 7, §4.1) are self-vindicating in mechanism and established in grade, since a mechanism cannot deliver more security than the premise it runs through. The partition's exhaustiveness (Chapter 3) is true by logical form; the structural dependence thesis (Chapter 2) is established by cross-theoretical convergence; the termini (Chapters 4–6) are established by argument a determined critic might challenge, the temporal terminus carrying one brute fact about obtaining. No result is more secure than the weakest premise it inherits.

6 An Auditable Chain of Dependencies

The book proceeds by derivation from a single foundational commitment, with each chapter's results conditional on those before it. A chain built that way can be audited, and the Premise Register is the audit; its rows are grouped under the claims of §3, so that it is that table expanded. It lists, result by result, what each rests on and what grade it carries, so that a reader who rejects a premise can see at once which results fall with it and which stand untouched. The compounded requirements of the foundational sequence — which results hang on no branch, which are branch-relative, and which questions stand deferred — are set out in the Register under Compounding conditions, where the full chain can be followed.

The purpose is that the cost of disagreement should be locatable. A reader need not accept or reject the book whole: refusing one premise removes a determinable set of results and leaves the rest standing, and the Register says which set. Two features of the chain follow. The first is that a premise's weight is not its position. Some early results carry little downstream, while a few later ones carry whole sequences, and §9 names the two that carry most. The second is that dependence is often partial. A later result may rest on a strong form of an earlier premise for

one of its conclusions and on a weaker form for the others, so that a critic who grants only the weaker form keeps most of what follows and loses one identifiable verdict. The clearest case of this shape concerns the account of value in Chapter 14, and the Register records the split.

Not every premise the book uses is derived. A small number are contested: positions the book adopts, argues for where it can, and, where it cannot, states what refusing them costs. The Register lists them under Contested premises, with what each costs to refuse and which rows fall with it, and the reader meets each by name at the chapter that first uses it. The same holds for the framework's limits. Where the account returns no further fact, the chapter concerned says so, and the General Conclusion (§3) gathers those residues in one place.

7 Where the Book Stands

The method differs from comparable projects in one specific respect. Parfit, Sider, and Schaffer each begin with a substantive philosophical position and defend it against alternatives. The present book begins with a premise whose denial exhibits it — at least one difference obtains — and derives consequences from it, adopting along the way a small number of premises that are contested and named as such (§6; the Register's Contested premises). The structural dependence thesis, which the derivation reaches immediately, is itself a substantive commitment, though one about content as such, confirmed across competing semantic frameworks. The premises are minimal and meant to be difficult to deny; the interest lies in the surprising richness of the consequences.

Parfit's *On What Matters* (2011) is the closest analogue in scope. Parfit argues that the major ethical traditions converge on the same fundamental principles; the book shares the conviction of convergence but offers a different diagnosis: what the traditions share is a common structural defect, the great moral conflation. Sider's *Writing the Book of the World* (2011) develops a systematic metaphysics from a foundational commitment to structure; the present book locates the foundational structure at a different level: at the condition, distinction,

that makes jointing possible, not at the joints of fundamental physics. Schaffer's priority monism holds that the cosmos as a whole is the fundamental entity; the book's grounding limitarianism reaches the opposite terminus: equality, a state of zero complexity, where Schaffer places a maximally complex whole. The disagreement is engaged directly in Chapter 4.

8 What the Book Does Not Attempt

The ambition is architectural, which is also where it is vulnerable. A specialist in any domain will find passages that move too quickly. This is the cost of breadth, accepted openly: the book is a single argument across twenty-four chapters, its strength — the cross-domain illumination of one structural vocabulary — inseparable from its weakness, the depth a dedicated monograph would provide. Whether the trade-off is worthwhile is for the reader to judge.

Beyond that, the book does not solve the hard problem of consciousness, offer a complete political theory, or provide a comprehensive theology. Chapter 17 distinguishes perfection (complete alignment with a standard on available evidence) from infallibility (the impossibility of error); the book applies the distinction to itself. Its foundational result — that at least one difference obtains — is infallible by logical self-vindication: the denial instantiates what the result asserts, on two premises of the same standing. Everything derived from it is at best perfect (the best achievable on the available premises) and revisable because those premises can be challenged. The book practises what it preaches: its foundation is logically secure; its superstructure is revisable. The Register's own epistemology, why the self-vindicating tier alone is infallible, is given at Chapter 17 (§4.3).

9 Where to Start Reading

Three routes follow the arch of §3, which Figure 1 draws: the left pier and its curve for the first, the right pier and its curve for the second, the whole structure for the third. A reader whose interest is the footing and the theological side can stop at Chapter 12 with a complete

argument. A reader whose interest is the metaethical side — mind, value, and ethics — can begin at Chapter 13, taking Chapters 2 and 6 on trust: the structural sequence needs only the determination dilemma (Chapter 2, §2.4) and the zero-complexity constraint (Chapter 6, §2). A reader whose interest is peace can begin at Chapter 14, since the political chapters inherit the metaethical pier and the theological side's clearing of the wall, and read the rest as the derivation of what those chapters assume; the keystone itself (Chapter 24) rests on the theological pier as well, so that reader adds Chapters 8–12 before the last chapter. Appendix A can be read first: it assembles the book's answer to its oldest question from results established in Chapters 2–12 and adds none of its own. Whichever route is taken, the Premise Register is the place to test a stone before standing on it, since it names what every result rests on.

Two results carry disproportionate load, and two constructed premises sit at the joints of the arch, so a reader deciding where to press should know them in advance. The determination dilemma (Chapter 2, §2.4) is the premise beneath the vindication law, the closure of the space of meaning, and the metaethical trilemma; the response-constitutive account of value (Chapter 14) is the premise beneath the moral sequence; its tracking floor, held by realists about value too, is what the purposive and political sequences use, as the Register's rows record. The General Conclusion (§2) marks both as the critic's shortest routes in and names the two joints beside them — the containment fork (Chapter 10, §3.2) and the premise that coupling is binary (Chapter 13, §2.3); the Register grades everything downstream of each.

10 What Volume I Established

This volume rests on the first. Volume I began from the one claim that cannot coherently be denied, that at least one difference obtains, and built on it the footing and the theological side of the arch: the first twelve of the twenty-four claims in the table of §3, each at the grade §5 assigns. No result established there rests on anything here, and every result of this volume rests, through the chain the Premise Register records, on results established there. The Register is printed in both

volumes for that reason, and each chapter's Presupposes line names the results it inherits.

The metaethical side needs less of that footing than its length suggests. The structural sequence of Chapters 13–16 uses the determination dilemma (Chapter 2, §2.4) and the zero-complexity constraint (Chapter 6, §2), and the sequence can be entered at Chapter 13 with those two taken on trust. The political chapters inherit the metaethical pier together with the theological side's clearing of the wall: that no mind stands above the hierarchy of difference (Chapters 8–12). The keystone rests on the theological pier as well, since the companion God of Chapter 24 is what survives the critique of Chapters 8–12, so that chapter is read with the pier in hand.

What Volume I removed matters here as much as what it established. It took away the authority a creed appeals to, a mind above the hierarchy that could settle a dispute from outside, and it set up no authority in its place. With that authority gone, the question of what could settle a dispute at all is the question the metaethical side answers, and peace, the keystone, is where the two sides arrive together. The General Conclusion (§1) gives the verdict on each of Volume I's claims, and the faults its §2 marks are where a critic of this volume should also look first.

11 An Invitation

Chapter 21 argues that the only truly stable resolution of disagreement is deep persuasion: the transformation of preferences through engagement, not their suppression through force. The book's own relationship to its reader takes the same form. It offers a perspective and invites the reader to test it against the derivation itself.

If the argument succeeds, it succeeds because the reader recognises something in it: the relationships it identifies correspond to relationships she can verify. If it fails, it fails because she identifies a premise that does not hold, an inference that does not follow, or a consequence the framework has not earned. A demonstrated flaw is more valuable than an unexamined agreement, since it specifies where the argument must

be strengthened or abandoned. The framework's own fallibilism ensures this: a book that argues perfection does not entail infallibility cannot exempt its own superstructure from revision.

The book opened with a question about conflict, and answered it by starting from the one thing that cannot be denied. Volume I carried the answer to the ground; what follows from it on the side of value and peace is the subject of the twelve chapters of this volume.

CHAPTER 13

Consciousness as Differentiation

Presupposes: the structural dependence of content on contrast (Vol. I, Ch. 2); the zero-complexity constraint (Ch. 6).

Establishes: the categorical transition principle, and the hierarchy derived from it level by level; the two axes of the mental, and the directional difference as the genus of directive response; the reinterpretation of four rival theories of consciousness, two of them in the appendix; and, in the appendix, the uniqueness principle, which no later result uses. All four results are graded constructed; note 1 records the grading structure.[1]

In brief. A system can discriminate a million shades of colour and be no more advanced, in the relevant sense, than one discriminating two. What raises a mind's level is not more distinctions but a categorically new *kind* of distinction — and this chapter derives the ladder of kinds, from thermostat to moral agent, from a single principle: content requires contrast. Along the way it separates the two axes every mental state divides into — *content* (what is detected) and *direction* (the felt pull toward or away) — and locates the boundary of experience at availability to thought, and the boundary of *mattering* at valence. It also names the question every theory of consciousness must answer before the hard problem arises — which physical events constitute experience, and what distinguishes them from the billions that do not? — and answers it in an appendix, with a proposal the rest of the book does not use. The chapter does not solve the hard problem as it is posed; it offers in its place an identity — feltness as availability to thought — that a critic can doubt, explains why the doubt arises, and constrains any solution.

---

1 Introduction

What makes a thermostat different from a thinking mind? Not the presence of causal sensitivity: both respond to their environments. Not the number of distinctions registered: a thermostat with a billion sensors would still not think. What makes the difference is the *architecture* of the distinctions detected: the kinds of contrast a system is sensitive to, and how those kinds are related to one another.

Chapter 6 established the central negative result: a state of zero complexity, with no internal distinction of any kind, excludes cognition, representation, and experience. This chapter develops the positive counterpart, extending the foundational commitment that content requires contrast into a theory of consciousness.

It asks what structural conditions a system must satisfy to possess conscious content, what explains the qualitative differences between levels of mental complexity, and the question every competing theory must answer: what distinguishes the physical events that constitute consciousness from those that do not? This is the boundary question, which is prior to the hard problem.

The chapter states conditions that any theory of consciousness must respect. It does not solve the hard problem as Chalmers (1996) poses it, why there is experience at all. In its place it offers an identity, feltness as availability to thought, as a claim a critic can doubt, with an account of why the identity seems false to the mind that holds it; and it constrains any rival solution. It keeps its own answer to the boundary question apart from what it derives, in an appendix on which no later result depends.

The chapter's positive result is a ladder: the levels of mind, from the thermostat to the moral agent, each added by a new kind of distinction a system can detect. Section 2 derives it, saying what makes each step a new kind, and with it the two axes every mental state divides into, content and direction. Section 3 places the rival theories against the ladder, §4 states what would refute it, and §5 meets the objections. Section 6 closes with what the rest of the book takes from the chapter, which is less than the chapter. The appendix (§§A.1–A.8) holds the chapter's own proposal for where experience is located and why, with the rival theories and the test that bear on that proposal. The ladder itself comes first, in miniature, a line and an illustration to each level.

Level	**What is newly told apart**	**An illustration**
0 Registration	A difference is reacted to, and remains unfelt	The thermostat's click; the reflex arc's withdrawal
1 Sentience	A registered difference becomes available to thought, and so is felt	Warmth in a hand, felt whether or not it is thought about

Level	What is newly told apart	An illustration
2 Simple consciousness	What is present to the senses is told apart from what the mind generates	Seeing red against imagining red
3 Embodied feeling	The body that feels is told apart from bodies merely perceived	The hand that hurts when pinched, against the hands only seen
4 Self-consciousness	The feeling body and the thinking mind taken as one thing, against all else	Pain in this hand and worry about it belong to one self, before either is reflected on
4+ Somatic integration	The one self extended to its parts that do not feel	Hair and bone counted as one's own, though unfelt
5 Other-consciousness	Another body is detected as a subject, where before it was an object	Another's pain registered as pain, which is empathy; the same capacity run in reverse is cruelty
6 Reflection	The system's own states become objects it can detect	Noticing that one is angry
7 Moral consciousness	One's own conduct is evaluated by what it does to other subjects	"What I did hurt her, and that counts against it"

2 The Categorical Transition Principle

2.1 Detection and availability

A thermostat responds differentially to temperature, yet nothing about the response is felt. A nervous system responds differentially to temperature; at some point the response becomes an experience of

warmth or cold. The difference is not the number of states discriminated, nor the fineness of discrimination: a thermostat can be made arbitrarily fine-grained without becoming a subject. The difference is that in the one case a distinction is merely registered, while in the other it is registered *for the system*: undergone, where the thermostat's response merely occurs.

What "registered for the system" comes to can be stated in full. The distinction is available to thought — the kind of content the percept/thought system can take up, that system being characterised structurally, without phenomenal vocabulary, at §5.2 — whether or not any thought is in fact directed upon it. Availability is a standing property of the state; nothing has to fire for it to hold. A felt warmth is available to be thought about, and is felt whether or not it ever is.

The thermostat's registration has no such availability: it is sealed in the reflex arc, graspable by no thought even in principle. Blindsight is the instructive intermediate (Weiskrantz, 1986). The information is registered and drives behaviour, yet it is unavailable to the patient's thought, and for that reason unfelt. What divides the registered-for-the-system from the merely registered is availability in principle; the exercise of thought is not required.

The foundational commitment fixes where the line falls, because content requires contrast. A bare causal transition carries no content; it is a change, not a difference for anything. Registration becomes experience where the registered contrast is available to thought. Within experience, a registered contrast may further acquire a direction, the system drawn toward one pole and away from the other.

This directional pull is valence in its rawest form. Present before any actual thought and attached to no particular thought, it is the most primitive ancestor of emotion and the boundary of mattering. Where an available contrast is felt as a pull, the subject has a stake in it. Where available but uncoupled, it registers without caring. Where not available at all, there is only a mechanism.

2.2 The principle stated

Quantitative enrichment is not advance. A system can discriminate a million shades of colour and be no more advanced, in the relevant sense, than one discriminating two; it has accumulated more distinctions within a single category and crossed into no new one. A qualitative advance is the detection of a categorically new kind of distinction — new when its obtaining presupposes the items the prior categories make available while itself falling outside every prior category.

The principle. This is the categorical transition principle: increases in mental complexity are constituted by the detection of categorically new kinds of distinction, while quantitative increase within an existing category leaves the level unchanged. A distinction can be categorically new in one of three ways, the third a limiting case; the hierarchy uses all three.

The first is a new dimension: a contrast on an axis that did not previously exist. The percept/thought distinction is categorically new relative to red-versus-blue, because it contrasts two kinds of mental state, where red-versus-blue contrasts two instances within one kind. The second is a new composition: existing relata integrated into a unit that affords a contrast none of its components affords. The self/non-self distinction is not the conjunction “has sensation and has thought” but the recognition that the sensing thing and the thinking thing are one entity, which can then be set against everything it is not. The unity is a relatum that did not exist before the integration.

A limiting third form is a new order, in which the system takes its own states as objects, so detection itself becomes something detected. Marking the form of each transition is what makes the hierarchy a derivation, where a list would leave each step unmarked: at every step it can be said which kind of advance has occurred, and why it qualifies as categorical.

2.3 Two axes: content and direction

The content axis. The progression tracks two things that must be held apart, since conflating them is the source of several persistent

confusions. The first is content: the kinds of distinction a system can detect or compose, which is what the numbered levels enumerate.

The direction axis. The second is direction: the pull toward or away, introduced at the first conscious level. Direction is not a level. It is valence, the felt toward-or-away, switched on with sentience and coupling to whatever content the higher levels make available. A thermostat and a sweating body both register heat unfelt, without valence, whatever electrical activity they carry. The levels build the content; valence runs through it.

What the separation fixes. This separates two categories of feeling that ordinary usage runs together. A sensation is localised — the warmth in a hand — and that localisation is part of its work, since it is what lets felt sensation demarcate the body. An emotion is not localised but a response of the organism as a whole, carrying a direction. Sensation answers what and where; emotion answers better or worse for the organism as a whole. Pain displays the composition coming apart: its nociceptive component is localised, so one can say where it hurts, while the suffering that gives it its negativity is not. The two dissociate, the sensation surviving intact when the suffering is removed.

The two axes are exhaustive; the exhaustiveness divides into two tiers whose grades differ. That every feature of a mental state is a determination or nothing is derived, and carries the derivation's strength (Chapter 6, §2.2). That a determination either carries a felt orientation or does not, and that orientation is the only further division a determination admits, is this chapter's own claim, carried at its constructed tier; the three cells — determination, orientation of one, or nothing — are what the derived partition yields once that claim is granted. Which cell a contested feature occupies, and what its occupancy consists in, can require substantive identification: the placement of feltness, below, is the availability thesis's work, and carries its register. The result, the two tiers together, is one later chapters lean on and Chapter 14 (§3.5) re-derives independently for its metaethical use.

The exhaustiveness argument. The derivation runs from a single question. Consider any feature of a mental state whatever. Either the

feature is a way the state is, a determination, one way rather than another, or it is not. If it is, it is content in the structural sense of Chapter 2: a diff, individuated by what it excludes, belonging to the first axis. If it is not — if the feature is no way the state is — then it is not a feature at all, by the dilemma of Chapter 6 (§2.2).

What remains to ask of any content is a single binary question: is it carried by valence or not, felt as a pull toward or away, or merely registered? Coupling is binary, and valence is not a further content but the orientation of one. There is therefore nothing for a third axis to be. A proposed third feature of the mental must be a determination (the first axis), the orientation of a determination (the second), or no way the state is at all, which is nothing.

One inheritance must be refused. The exhaustiveness does not descend from the partition of Chapter 3; the temptation to say it does should be resisted all the same, since the two divisions do not correspond. Chapter 3 divides on distinction or its absence, which yields determination or nothing — two cells, with no content/direction split among them. The trichotomy here is a sub-division of the first cell: a determination either carries valence or does not, and "the orientation of a determination" is a determination under a description, not a third thing beside determinations.

What the partition supplies is the exclusion of the third cell, since a feature that is no way the state is at all is nothing. What supplies the rest is the substantive claim that coupling is binary and that valence is the only sub-division a feature of the mental can take, which is carried at this chapter's tier. Chapter 14 (§3.5) does not rest on it. Its trilemma is re-derived there on a division of its own — transferable or not — so a reader who declines the two axes may still grant it.

One objection tests the partition where it is weakest, and meeting it fixes the sense in which the axes are complete. The objection points to subjectivity itself — that an experience is undergone, that there is something it is like, that the state is someone's — and claims this is a third feature, neither a content nor the orientation of one. The reply

separates two things the objection runs together: that experience is felt at all, and that it is felt as belonging to a self.

Take feltness first. The partition by itself forces only that a residue which is *no* way the state is at all is nothing (Chapter 6, §2.2). What it does not by itself deliver is the identification the account needs here: that feltness is *the availability of a content to thought* (§2.1). A felt state is content of the kind the percept/thought system can take up, whether or not it is taken up: availability in kind, an episode of access being a further matter. The unattended periphery of the overflow cases (Chapter 7, §2.3) is content the system can take up though it does not, which is what makes overflow a datum for the identification instead of a counterexample to it. Call this the **availability thesis**: a substantive identification, not a corollary of the partition, and carried in the Register as constructed. Granted, feltness is a property of the content and belongs to the first axis — a placement the partition guarantees is available, since no third cell exists for feltness to flee to, and which the thesis, not the partition, performs.

The residue and its two tests. The realist who resists presses a sharper residue: a determinate feature of available content, distinct from its availability — available content that could, she says, have been present unfelt. Against that residue the availability thesis has its reply, which should be weighed for what it is: a costing of the residue.

Two considerations support it. First, the residue must, by Chapter 2, carry a contrast. The realist owes a specification of what an available-but-unfelt content of the very same type would be, a single content held fixed with its feltness subtracted, a zombie drawn at the scale of a single state; no case exhibits one. Second, the equalisation evidence (Chapter 6) shows feltness and availability co-eliminating: strip a modality's contrasts and its felt field vanishes altogether, never thinning toward an unfelt-but-available residue. Identity predicts this; a further-determination view must record it as coincidence.

Neither consideration defeats the intuition that the residue is there; the chapter does not claim they do. What the projection mechanism of Chapter 8 explains is why a mind built from distinctions finds the

residue inevitable, which is an account of the intuition, short of a proof against it. One further instrument is available to a reader who accepts it elsewhere. Chapter 14 (§3.6) tests a contested clause by subtraction: remove it and ask what contrast remains. Remove "felt, over and above available" and nothing changes in any register, the subject's own reports included. By that test the residue is vacant, as the realist's clause is vacant there. A reader who refuses the test refuses it in both places at once, so the Register counts the refusal as one exit and not two.

The intuition that a fact is nonetheless left over is itself explained, on the projection pattern of Chapter 8 (§3.4). A mind whose exercise of its feltness-concept is itself a felt taking-up must find the identification incredible, since feltness is presented from the inside in a mode no structural description shares. The incredulity is therefore what the thesis predicts the realist would feel whether or not any further fact existed, so it cannot discriminate his view from the identity. What the literature calls the phenomenal concept strategy (Loar, 1990; Papineau, 2002) is here a consequence of the framework's own account of representation, with no auxiliary hypothesis needed.

The second half of subjectivity — that the experience is *mine* — is a content, and a specifiable one. Ownership is a composition: the system represents a body, represents itself, and represents the feeling-body as the seat of that self (Level 4, below); mineness is that composed representation, on the first axis by construction. This is why felt experience at the floor is undergone yet not owned: availability makes it felt before any self-model is present to make it personal. A sentient creature is already a subject (there is something it is like to be it) while representing no self for its experience to belong to.

What the axes settle. Neither half of subjectivity is a third axis. Feltness is availability of content; ownership is complex content; valence, where present, is the second axis, supplying mattering and not feltness. The realist's challenge is now definite: exhibit a single available content whose feltness can be subtracted while the content is held fixed, or a mineness that decomposed into no self/non-self contrast. Either would falsify the architecture here.

The same constraint binds Chapter 14's target in advance. A philosopher who posits a sui generis mode of the mental — a rational grasp that is neither the registration of structure nor the feeling of pull — must characterise it; every characterisation will attribute determinate character (the first axis), oriented character (the second), or no character, which is nothing.

The quarantine extends to the whole chain by a fork the realist is owed. Grant the phenomenal residue arguendo, a feltness over and above availability. Either it acts — affecting report, behaviour, or orientation — in which case it is a determination absorbed into content or direction. Or it idles, a pure spectator touching nothing, changing no downstream result, since value, morality, purpose, and persuasion all run on registered pulls. Every result from Chapter 14 onward holds with or without the glow, so the residue dispute, wherever it settles, is a local dispute about this chapter's positive theory and never a lien on the chain.

Because valence is separable from content, any content can be taken up in either of two modes. Coupled to valence, it is evaluated: felt as mattering, oriented as better or worse. Uncoupled, it is felt but neutral where available, and merely registered where not — computed, or checked against a standard, without mattering — orthotic processing, in the term Chapter 14 (§4.3) fixes. Call a content of the first kind, a difference carried by valence, a directional difference.

The vocabulary can now be fixed. Valence is the bare pull. An emotion is valence bound to content, a directional response about something, of which human moral emotion is the high-content case. The felt bindingness of a rule is a further, misattributed instance (Chapter 16). This single fork, recurring at every level, is the structural root of the distinction between evaluative and cognitive operation developed there: what appears in that chapter as a feature of moral judgement is a general feature of mind, present wherever content can be processed with or without valence engaged.

2.4 Structural containment

The levels are ordered by structural containment: a higher level presupposes and incorporates the categories the lower levels supply. A self-model (Level 4) cannot be assembled without the feeling-body and mind categories it integrates (Level 3). The recognition of others as subjects (Level 5) presupposes the self whose structure it extends. Damage to a lower capacity degrades the higher capacities that depend on it, never the reverse.

This is a substantive structural claim, since it carries a falsifiable prediction, which no stipulation does. The levels are separable in one direction only: a system may possess a lower level without the higher, but not a higher while wholly lacking the lower one it composes. A system shown to possess an integrated self-model while wholly lacking the percept/thought distinction would falsify the ordering. Containment is an ordering of structural dependence, not of temporal sequence, so the levels need not be acquired in order; a mature system runs many concurrently. The claim is only that the higher cannot obtain without the lower.

2.5 The hierarchy

With the principle and the two axes in place, the hierarchy can be set out. Each level names the categorically new distinction it adds, the form of transition it represents, and the coupling of valence it makes available. The table compresses the derivation; the notes after it record only what later chapters use.

Level	Content axis: the new distinction	Form of transition	Direction axis: valence
0 Registration	Differential causal response unavailable to thought (thermostat, reflex arc)	— (pre-conscious floor)	none; unfelt

Level	**Content axis: the new distinction**	**Form of transition**	**Direction axis: valence**
1 Sentience	Registration become available to thought, and therefore felt	availability crosses mechanism → experience	valence may couple: neutral if not, directional if so
2 Simple consciousness	Percept vs thought — the present-to-sense vs the internally generated	new dimension	couples to thought: one representation = instinct; linked representations under one direction = reasoning; uncoupled = orthotic computation
3 Embodied feeling	Feeling-body vs merely-seen body; thought recognised as a category	new dimensions	localised sensation marks the body as the system's own; valence merely couples more or less strongly
4 Self-consciousness	The feeling body and the thinking mind are one entity, set against the non-self	new composition	valence becomes personal — the pull is owned
4+ Somatic integration (parallel branch)	Connected-non-feeling (bone, hair) vs unconnected-non-feeling	new composition, branching	— (no later level depends on it)

Level	Content axis: the new distinction	Form of transition	Direction axis: valence
5 Other-consciousness	Other bodies are subjects	new composition, by extension	valence takes others' feelings as object = empathy; reversed = cruelty (one capacity, two directions); uncoupled = cold theory of mind
6 Reflection	The system's own states become objects	new order	—
7 Moral consciousness	Empathic evaluation (5) brought to bear on the system's own conduct, taken as object (6)	new composition	concordant direction = moral goodness; reverse = cruelty toward an end

Three notes carry forward.

At Level 1, availability is the on-switch of experience and valence is the on-switch of mattering. One dependency in the ordering should be recorded here. Availability is availability in kind to the percept/thought system; that system's contrast with the percept is the dimension Level 2 introduces, so the first level is characterised by reference to a format whose full articulation arrives one level up. The ordering is one of containment; a format can be characterised before the contrast that displays it is detected (§5.2); the relation is a dependency to be read in that direction, not a circle. Nor is a Level 1 creature felt by courtesy of a consumer it lacks. The consumer is present at Level 1: an internally generated system that takes registrations up is what makes them available in kind, and the three marks of §5.2 fix it without any mention of feltness. What Level 1 lacks, and Level 2 adds, is the detection of the

contrast between what that system takes up and what it generates. Availability at Level 1 is availability to a system the creature has, whose difference from the percept it does not yet register.

A sensation available to thought but carrying no pull is felt yet neutral — a grey field, a bare touch — so the account affirms neutral felt experience as the ordinary case and owes no strained treatment of it. That valence, and not fineness of discrimination, is what turns a felt sensation into an emotion is fixed here.

At Level 5, empathy and cruelty are one capacity run in opposite directions, with no second mechanism, which is the result Chapter 16 (§6.3) redeploys against the persecutor. The cold-theory-of-mind case is Level 5 content processed without valence, which the analysis of psychopathy requires (Chapter 16, §5.2).

At Level 7, that valence constitutes evaluative difference, and does not register values obtaining independently of it, is argued in Chapter 14; the present chapter places the capacity, borrows the name “moral” from the identification Chapter 16 argues, and leaves its metaphysical analysis to that argument. The progression from 0 to 7 is generated by a single principle, content requires contrast, applied through the two forms of categorical transition and threaded by a single valence. Each step is a determinate advance over the one below, its place fixed by what it presupposes. A being that detected every category the principle permits would instantiate every level at once.

The derivation the table compresses can be run in full, one level at a time, by the test §2.2 states: does the level’s distinction presuppose the items the prior categories make available, while itself falling outside every prior category? At each level the test returns the same verdict, so the form of each transition can be read off. The levels themselves follow.

Level	The new contrast	What it presupposes, and why no lower level draws it	Form
2 Simple consciousness	The present-to-sense against the internally generated	Felt registration, since both relata are felt; Level 1 contrasts registrations within one kind, never two kinds of mental state	New dimension
3 Embodied feeling	The body that feels against bodies merely seen; thought marked as a category	The percept/thought dimension: only a system that separates the two can locate a sensation in one perceived body among others, and recognise the internally generated as a kind	Two new dimensions
4 Self-consciousness	The feeling body and the thinking mind as one entity, set against everything it is not	Both relata; the unity is a relatum that did not exist before the integration, and the self/non-self contrast is available to nothing below	New composition
4+ Somatic integration	The unity extended to parts that do not feel — bone, hair — against unconnected non-feeling things	The self-model; no later level depends on it, which is why it branches	New composition, branching

Level	The new contrast	What it presupposes, and why no lower level draws it	Form
5 Other-consciousness	Another body detected as a subject	A self to extend; a subject other than oneself is a relatum no prior level affords	New composition, by extension
6 Reflection	The system's own states as objects	A self whose states they are; the contrast between a state and that state as an object is of a different order from every contrast below	New order
7 Moral consciousness	Empathic evaluation brought to bear on chosen conduct	Levels 5 and 6; the object — the system's own conduct as it bears on other subjects — is a composite relatum neither affords alone	New composition

Two natural candidates test the count and not the ordering; running the criterion on them shows why the hierarchy derives its levels instead of listing them. Episodic memory looks level-like, since a mind that re-lives its own past seems transformed in kind. Apply the test: does its obtaining presuppose the prior categories while itself falling outside every one of them? Episodic recollection re-presents contents of kinds already on the table — percepts, feelings, the self who had them — ordered in time, and temporal ordering is detected wherever change is detected, from the lowest levels up. Its distinctive signature, ownership of a past experience, is the Level 4 self-model applied along a dimension it already carries, which makes episodic memory a

spectacular quantitative enrichment on the wrong side of the quantitative/categorical line.

Language invites the same verdict by a different route. Its representational content rides on distinctions the prior levels supply; its new contribution, the public labelling act, is what Chapter 7 (§3.1) files as the third grid's work — a loop within the grids, adding no new kind of detected distinction. What language transforms is transmission and combination; what a level requires is a new category of detectable difference. Both candidates fail the criterion for stateable reasons; the stateability is the point: a critic who thinks a level is missing knows what she must exhibit, a kind of distinction that presupposes the existing furniture and falls outside every category on the table.

A third candidate is agency, the chosen against the merely occurring; the criterion turns it away for a reason that bears on the whole hierarchy. Every level is a level of registration: what the system detects. Acting is not a further thing detected. It is what the system does on what it detects; it is present from the floor. The thermostat's click and the reflex arc's withdrawal are output conditioned on registration unavailable to thought; the shout of fear is output conditioned on a felt registration; a choice is output conditioned on alternatives the field has registered, which the percept/thought distinction of Level 2 first supplies, an internally generated course set beside the perceived one; deliberation is output conditioned on alternatives and on the system's own states taken as objects, which is Level 6. What changes across the hierarchy is what the output is conditioned on; the complexity of the act scales with the complexity of the field that feeds it, so "choice" names the act at the first level where alternatives are in the field and adds no category of detected difference of its own. The rule that turned away episodic memory turns away agency: the chosen/merely-occurring contrast is composed from the alternatives Level 2 first supplies and the reflection of Level 6; a system with both has the contrast without a further level. The criterion is registration, not anatomy — a spinal reflex is central nervous tissue; what places it at the floor is that nothing in it is available to thought.

Level	Output conditioned on	The act, named
0	differential causal response unavailable to thought	The reflex, on an unfelt registration
1	registration available to thought, and therefore felt	Felt reaction: the shout of fear, on the felt registration alone
2	the present-to-sense distinguished from the internally generated: the first alternatives in the field	Instinct, where valence couples to one representation; choice, where representations are linked and the act runs through the alternatives
3	the body registered as the system's own	A choice about the body registered as one's own
4	the feeling body and the thinking mind as one entity, set against the non-self	A choice for oneself, the options held as a self's own
5	other bodies registered as subjects, so that their perspectives are among the alternatives	A choice conditioned on another's perspective: the threshold of sufficiently complex choice (Chapters 14 and 16)
6	the system's own states taken as objects	Deliberation: the alternatives, and one's own states taken as objects
7	the system's own conduct as it bears on other subjects	Moral choice: one's own conduct, evaluated by what it does to other subjects

The output axis is also the axis of choice; instinct is a route on it and no rung. The instinctive act is the one conditioned on the felt registration alone, bypassing whatever alternatives the field contains; the route stays open at every level from the first: a dog bites on instinct though its field has alternatives, a person flinches though hers has more. Choice is the act that runs through the alternatives; its complexity scales with the field that feeds it. What Chapters 14 and 16 call a sufficiently complex choice has its address here: an act that runs through the alternatives and not by the instinctive route, in a field in which the affected party is registered as a subject (Level 5), with reflection (Level 6) supplying the agent's own conduct as an object for the evaluation Level 7 performs.

One thing the output axis does carry, and Chapter 18 draws on it. What a system does is never among the things it registers: the producing of an act is not a content of the field that produced it, at any level. Where the act is choice, this is the asymmetry a chooser meets from inside.

Where the argument stands: the categorical transition principle and its hierarchy are derived (§2), on the two axes and the availability thesis; they are what the chapters after this one draw. The boundary question and the chapter's own answer to it — the uniqueness principle, the interface, and the equalisation constraint — are set out in the appendix (§§A.1–A.8), where the chain does not use them. The rival theories can now be placed against the hierarchy.

3 Engagement with the Literature

The rival theories of consciousness stand in determinate relations to the hierarchy: two are complemented, the nearest neighbour of the availability thesis is placed, illusionism is met; the affective-neuroscience convergence corroborates the valence claim. The interface account's own rivals, integrated information theory and predictive processing, are met in the appendix (§A.6).

Higher-Order Thought theory (Rosenthal, 2005) holds that a state is conscious when it is the object of a suitable higher-order thought. The

account reinterprets HOT as a theory of one categorical transition, the transition to reflective self-consciousness (Level 6). It captures that transition correctly while misidentifying its scope, since Level 1 registration involves phenomenal character without higher-order awareness; the hierarchy supplies the levels below and above the HOT transition that the theory lacks.

Global Workspace theory (Dehaene & Changeux, 2011) holds that consciousness arises when information is broadcast globally, becoming available for report and control. The categorical transition principle complements it by supplying what is accessed. GWT identifies the mechanism of access but not what makes the broadcast content conscious as well as accessible: the content is conscious when it involves the detection of categorically distinct contrast, which the broadcast then makes available for integration.

The nearest neighbour of the availability thesis in this literature is Prinz's (2012) account, on which a representation is conscious when attention makes an intermediate-level perceptual representation available to working memory. The two agree that a consumer fixes the boundary of experience and differ on the consumer — attention to the intermediate level there, availability in kind to the percept/thought system here — and the difference is the one §A.7 stakes on the phosphene case. Tye's (1995) poisedness condition is the older ancestor: a content is phenomenal when poised for use by the cognitive system, whether or not used. The availability thesis keeps the poise and drops the abstractness and nonconceptuality Tye required of the content, since the format the percept/thought system consumes is fixed structurally at §5.2 and by no further mark.

Illusionism (Frankish, 2016) holds that phenomenal properties do not exist and that a theory of consciousness owes only an account of why introspection represents them; the meta-problem programme (Chalmers, 2018) makes the second demand general, asking every theory to explain the intuition of a further fact. The availability thesis is the realist neighbour of the first and an instance of the second. It keeps feltness, identified with availability in kind, and denies only the residue; the projection mechanism of Chapter 8 is its answer to the meta-

problem, explaining why a mind built from distinctions finds the residue inevitable (§2.3). Illusionism and the thesis share the debunking and part on what survives it, the illusionist retaining the seeming and the thesis the availability; what tells them apart is whether the equalisation evidence, which co-eliminates feltness and contrast, is read as a fact about experience or about reports of it.

Affective neuroscience corroborates the chapter's most distinctive structural claim, valence, from a different quarter. The account holds that what converts registration into the lowest grade of experience is the switching-on of valence at sentience (§2.3). Panksepp (1998) reaches a structurally parallel claim from lesion and stimulation evidence: the primordial form of consciousness is affective, phylogenetically prior to the cortical machinery of perception and thought, with cognition layered above it. That is the ordering the two-axes architecture predicts, since valence is the first thing that makes a registered difference a difference for the system.

The convergence is undesigned. Panksepp proceeds from the neuroscience, the present account from the structural dependence of content on contrast; the two reach the same ordering. One divergence remains: Panksepp treats affect as a natural kind delivered by specific neural systems, while the account treats valence as a structural role those systems realise in the mammalian case without exhausting, since the exhaustiveness argument of §2.3 makes valence a feature any possible sentience must instantiate, whatever its substrate.

4 Falsification Conditions

The categorical transition principle generates predictions that could falsify it. Level 4 (self/non-self) without Level 2 (percept/thought) would fell the containment ordering. A qualitative increase in complexity through quantitative accumulation within one category, with no categorical transition, would fell the principle. A Level 0 system possessing any consciousness would fell the zero-complexity constraint. The availability thesis adds a condition of its own, stated at §5.2: a content shown to be phenomenal while unavailable in kind, of a format

the percept/thought system could not take up even in principle, would refute the identification. The interface account adds conditions of its own, stated with the stake they carry in the appendix (§A.7).

5 Objections and Replies

5.1 The hard problem objection

Objection. The chapter identifies structural necessary conditions without explaining why physical processes give rise to subjective experience at all.

Reply. Conceded as the objection poses it: the chapter offers no account of why there is experience at all. What it offers in place of one is an identity — feltness as availability to thought (§§2.1, 5.2) — which a critic can doubt, the residue of that doubt being the hard problem in the form it keeps here. Beyond the identity, the claim is that any solution must be compatible with the structural conditions identified here, and that the equalisation constraint (§A.5) narrows the space of viable solutions by ruling out theories on which experience and contrast separate. The residual sense of a gap, the explanatory gap of Levine (1983), is itself addressed at §5.2: it may be the representational limitation of Chapter 8, since a mind constituted by distinctions cannot picture the identity it theoretically endorses.

5.2 The access/phenomenal objection

Objection. Block (1995) distinguishes access consciousness, a state's availability for reasoning, report, and control, from phenomenal consciousness, there being something it is like to be in the state, and argues the two are dissociable. The availability thesis identifies feltness with availability to thought, so it appears to define the phenomenal into a species of access — the very conflation Block's distinction exposes. A circularity compounds the charge: "availability to thought" explains consciousness by appeal to thought, itself a conscious kind.

Reply. The identification is drawn at a different joint from the one Block targets; the difference is the one §2.3 marked: availability in kind

versus episode of access. Block's access is episodic and functional, so a state is access-conscious when its content is in fact poised for use. The availability thesis concerns a standing property: content of the type the percept/thought system can take up, whether or not it is taken up.

The dissociation evidence then falls into place on both sides of the thesis. Overflow — the unattended periphery, the uncued rows of the partial-report grid (Sperling, 1960; Block, 2007) — is phenomenality without episodic access, which Block adduces against identifying the phenomenal with access. On the availability-in-kind reading it is a datum for the thesis: the periphery is felt because it is available in kind; its going unretrieved does not make it unfelt. Blindsight is the converse, registration from which availability in kind has been removed, and with it the feltness. What Block's cases show is that feltness is not episodic access, which the thesis never claimed. A critic holding that they show feltness floats free of availability in kind as well owes the state §2.3 demanded: a single content of the very same type, available in kind, with its feltness subtracted. No dissociation exhibits it.

Availability is categorical, not dispositional. A dispositionalist misreading must be blocked. Availability in kind is not a bare disposition to be accessed but a categorical property of the state: its content is of the representational format the percept/thought system consumes, individuated by the contrastive structure of Chapter 2 and delivered at a terminus (§A.3). The state possesses that format whether or not the consuming capacity is free. Overflow dissociates the two grounds cleanly, since the working-memory bottleneck limits the consumer while the periphery's contents keep their format throughout; their feltness tracks the format and survives the bottleneck.

The circularity charge mistakes the register in which "thought" enters. What the thesis invokes is a structural role — the internally generated, inferentially linked representational system whose contrast with the percept constitutes Level 2 — and the role can be specified, and need not be gestured at, in three marks, none of which mentions feltness.

Origin: the system's states arise internally, from no sensory terminus, an architectural fact statable in the pathway vocabulary of §A.3, where origins and termini are already distinguished without phenomenal remainder. Linkage: its states compose inferentially, each content constraining and constrained by others in chains, a relational property of contents individuated contrastively (Chapter 2). Format: its states consume and produce a common representational format, individuated the same way.

Each mark is attributable from the third person, to a system about which the phenomenal question is still open. The analysis is therefore recursive but grounded, as a recursive definition is grounded by its base clause. Felt states are those available to a system some of whose states are themselves felt, while the system is fixed by the three marks and not by its feltness, so the definition can be entered from outside.

The discharge exposes where the thesis's substance lies and conceals none of it. With both relata structurally specified, the identification between availability so defined and feltness is a substantive claim a critic can coherently doubt, the residue of that doubt being the hard problem (§5.1); the Register's "constructed" records this (§2.3). What the discharge removes is the charge of circularity. What it leaves is the claim, standing where the Register always placed it.

Chalmers's (2007) dilemma for phenomenal-concept strategies is taken at its first horn. The dilemma runs: either the concepts are structurally explicable, and cannot explain our epistemic situation, or they are not, in which case the gap reopens. The projection mechanism of Chapter 8 is a structural account of representation; what it explains is not the truth of the identity but the inevitability of the intuition of its falsity, which is the epistemic situation to be accounted for. What is conceded is what §5.1 conceded: the demand for a further fact lapses only if the availability thesis holds, which the Register carries as constructed, its support and its limit stated where they are given (§2.3).

The empirical wing of Block's case has since been sharpened by the no-report paradigms (Tsuchiya, Wilke, Frässle, & Lamme, 2015); the refinement should be registered, because it bears on which theories the

dissociation evidence actually pressures. In these designs the contents of experience are tracked without asking the subject to report them — percepts in binocular rivalry read from reflexive eye movements, for example — and the neural correlates so isolated shrink, since much frontal activity previously credited to consciousness tracks the act of reporting instead.

The result presses hardest on accounts that tie feltness to access episodes, global broadcast for report included; the availability thesis stands on the other side of the cut it makes. Availability in kind is poise, never performance. The rivalry percept in a no-report design retains the format the percept/thought system consumes, since the subject can report it the moment she is asked; the reflex the decoder reads is downstream of the registered content; what the paradigm subtracts is the access episode the thesis never required. The paradigms also supply the falsification shape §2.3's challenge calls for: a content demonstrated phenomenal while unavailable in kind, of a format the percept/thought system could not take up even in principle, would refute the identification — and the no-report literature, for all its subtractions, exhibits none.

6 Conclusion

The chapter yields four results, all constructed, over an established constraint.

The categorical transition principle. Increases in mental complexity are constituted by the detection of categorically new kinds of distinction, ordered by structural containment; the hierarchy, Levels 0–7 with 4+ as a parallel somatic branch, provides the architecture. What generates the qualitative difference between a thermostat and a self-conscious being is a different architecture of the distinctions detected, not a different kind of relation.

The two-axes architecture and the availability thesis on which it turns are what Chapters 14–16 inherit — the two axes and the existence of valence, and nothing else from this chapter's positive theory. The inventory can be given by name, so the quarantine is checkable at a

glance. The trilemma of Chapter 14 (§3.5) draws the derived tier of the two-axes result — determination or nothing — together with a division of its own, transferable or not, so that it uses no result from this chapter's constructed tier. The anchoring of Chapter 14 (§6.4) draws the existence of valence as the orientation of a content. The grip argument of Chapter 16 (§3.1) draws the derived tier again, while the chapters beyond draw the modality's existence through those results. No downstream result draws the availability thesis in full, so a reader who rejects the placement of feltness loses this chapter's positive theory and nothing that later chapters use.

The uniqueness principle. What makes certain physical events constitute experience is their structural uniqueness at an interface with the modality's absence, identified through three ordered features and constrained by equalisation (Appendix, §§A.1–A.5). This rules out theories on which experience is intrinsic to particles, persists without contrast, or diminishes gradually. The panpsychist's proto-consciousness cannot answer the boundary question, because a property that marks no boundary can constitute no content.

The reinterpretation of rival theories. The hierarchy supplies the structure IIT lacks, reinterprets HOT as one categorical transition, complements GWT with the architecture of what is broadcast, contests predictive processing on a discriminating prediction (§A.6), and is corroborated on valence by affective neuroscience.

If consciousness is constitutively dependent on detected difference, the question becomes: how do conscious beings evaluate the differences they detect? Not all differences are alike — red-versus-blue is a chromatic difference; better-versus-worse is an evaluative one. What modality constitutes that last category? And what follows if it is not reason but emotion? Chapter 14 develops the answer.

Grades and Dependencies

1. **Grades.** All four results are *constructed* — positive proposals with live rivals — over the *established* zero-complexity constraint (Chapter 6). Two

gradings are internally structured and the Register keeps them apart: the two-axes exhaustiveness has a derived tier (that a feature is a determination or nothing) and a constructed tier (that a determination's only further division is by orientation, together with the availability thesis's placement of feltness), and the uniqueness principle separates its established boundary constraint (§A.2) from its constructed criteria (§A.3). The downstream-use inventory lives in §6 by design, where a critic can check the quarantine at a glance; the availability thesis is used by no later result; the Register marks it so. The neural vocabulary of §§A.2–A.4 is read realistically for convenience and carries no commitment the framework lacks: "terminus" and "interface" name stable patterns of registered distinction; every claim made in them survives translation into that idiom, the phosphene prediction included, which becomes the prediction that a registered signature meeting a registered absence yields a registration of the modality's kind (Chapter 1, §5.3).

Appendix: The Boundary of Experience

This appendix holds the chapter's own proposal for the boundary question, which no later result uses: the uniqueness principle and the interface at which it locates experience (§§A.1–A.4), the equalisation constraint any theory of consciousness must respect (§A.5), the two rival theories whose target is the proposal itself (§A.6), the predictions the proposal is staked on (§A.7), and the one objection aimed at it (§A.8). Its results are carried in the Register at their own tier; rejecting them costs the chain no result.

A.1 The boundary problem

Inside a human brain, billions of neural events occur every second, and most are not conscious. A motor command, a homeostatic signal, an immune-related response: each involves activation against surrounding non-activation, each has a physical boundary with its inactive neighbourhood, yet none constitutes experience. A pattern of activation in the visual cortex, triggered by photons transduced through the retinal pathway, is also a neural event — and this one is a conscious experience. What distinguishes it? Not anatomical location, which merely relocates the question. The question is structural; every competing theory must answer it. Physicalism gestures at "the right

organisation" without specifying what constitutes the crossing; panpsychism inherits the question at the system level (§A.5); dualism locates the boundary between substances without explaining interaction. Each offers a metaphysics; none characterises the boundary.

A.2 From the structural dependence thesis to uniqueness

The foundational principle answers directly. Content requires contrast, so any determinate content is individuated by at least one distinction (Chapter 2). A property shared by all physical events — mass, energy, charge, or "proto-consciousness" — cannot constitute the content of consciousness, because it is present on both sides of the boundary and therefore contrasts with nothing. Only features unique to the consciousness-constituting events can do the work.

A critic may charge that the structural dependence thesis concerns content individuation, so applying it to phenomenal experience equivocates. The charge would bite if the connection were stipulated. The equalisation principle (Chapter 6) supplies it as evidence instead: phenomenal experience and detected contrast co-vary as the thesis predicts, vanishing together categorically. The thesis applies to phenomenal content because phenomenal content behaves in accordance with it.

A.3 Three levels of uniqueness

The consciousness-constituting events are distinguished by three features, ordered by the kind of uniqueness each contributes.

Positional uniqueness: they are the terminal output of a dedicated processing pathway carrying a signal inward to a terminus, not restricted to external input, since visceral pain terminates in the insula, emotional experience in affective centres, proprioception in the somatosensory cortex. A carbon atom at the visual-cortex terminus and one in the kneecap are chemically identical; what differs is position. This is necessary, not sufficient.

Functional uniqueness: the events are the endpoint of a chain dedicated to carrying a specific input type inward, "pathway" meaning

the entire dedicated architecture, loops included, in place of a linear feedforward chain. Function narrows the field beyond position. A motor command is an origin, not a terminus; a homeostatic signal circulates regulatory information and carries no input type inward.

Structural uniqueness: the specific activation pattern, the triggered signature, that a dedicated architecture generates in response to a specific input type, defined by causal-structural role, whatever the physical composition. It preserves no copy of the input: what arrives at the higher visual areas bears little resemblance to the entering light, having been recoded through opponent processes, lateral inhibition, and cortical remapping. This is the deepest level and the one that constitutes qualitative character. Two events could share position and function yet, with different pathway architectures, produce different signatures — and the experience would differ.

The cochlear-implant case checks the inference against report. The implant transduces sound through an architecture unlike the intact cochlea's, and recipients describe a qualitatively unfamiliar mode of sound — a report the audiology literature attributes largely to the implant's coarse spectral resolution, which the account reads as one dimension of the altered signature: different architecture, different signature, different character, reported from the inside, with the inference from the second to the third the one the account stakes. The honeybee's tetrachromatic vision extends the same inference to where report is unavailable; there the different-character claim is a prediction the architecture licenses, with no datum the subject can confirm.

The order of discovery should be stated openly, because a reader will suspect it. The three criteria were found by asking what distinguishes the events that constitute experience from those that do not — reverse inference. The reply separates discovery from justification. What the framework derives is the constraint the criteria must satisfy: consciousness-constituting events must be distinguished by features unique to them (§A.2), since a shared feature contrasts with nothing.

Which uniqueness does the work is an inference to the best explanation of the dissociation evidence (§A.4); its warrant is

prospective, resting on the falsification conditions of §A.7, including the prediction that discriminates the interface account from its predictive-processing and global-workspace neighbours. The principle is carried as constructed, a positive proposal with live rivals, while the boundary constraint it answers to is established. The criteria could be wrong in a way the constraint could not, so the account stakes them on the predictions, whatever fit first suggested them.

A.4 The interface

The terminus activation occurs against the modality's absence: the condition obtaining when no activation of the relevant type is present — the blind field for vision, silence for hearing, numbness for affect. A distinction matters here. Darkness is a visual experience, minimal light detected against the visual field; the blind field is not a visual experience at all, since the congenitally blind person does not see darkness but has no visual phenomenology. The modality's absence is the structural zero of the detection system.

The experience-constituting events are located at a specific interface: the boundary where a terminus activation carrying a unique triggered signature meets the modality's absence. The quality question — why this character? — is answered by what differs across modalities, the signature, and not by what is shared, the absence and the particles. The absence is not a fourth ingredient. It is the contrast pole the signature needs in order to be a diff at all, which is Chapter 2's requirement applied at the interface. The condition holds of any mind.

One class of cases lets the interface reading do discriminating work beyond reinterpreting the conjuncts: those where a signature meets an absence that has just been created. After deafferentation the somatosensory terminus can fire into a limb that is gone, so the phantom is felt vividly at a location the body no longer occupies. The sudden silencing of a cochlear input can be followed by tinnitus.

On the interface account these are expected, since vividness follows from a signature meeting the modality's absence, whatever the peripheral source. The clinical picture is consistent with the reading and

does not select it, because accounts that locate the phantom in the reorganisation of the deafferented cortex (Ramachandran & Hirstein, 1998) predict persistence too; what it excludes is only a view on which experience tracks the intact input alone, which no live theory holds. That phantom and tinnitus are often more insistent than the veridical percept is what the interface reading expects; the test that separates it from its live rivals is the phosphene design of §A.7. Dreams are the complementary limit: internally generated activation acquires the characteristics of the live terminus signal and produces experience indistinguishable from perception, which the account predicts, since phenomenal character is constituted by the interface conditions, whatever the causal origin of the activation.

The harder test is the dissociation in which perceptual information guides behaviour while remaining unconscious; here the conjunction discriminates. In blindsight, damage to the primary visual cortex abolishes acknowledged visual experience across a region of the field, yet patients reliably detect and discriminate stimuli they deny seeing — residual capacity carried by routes that bypass the damaged area (Weiskrantz, 1986). This is the predicted profile. The residual pathway sustains a functional response, while the dedicated route that would generate the triggered signature at the visual terminus is the very thing the lesion removes; it is that conjunct, and not the bare registration, on which visual phenomenology depends. Blindsight is registration surviving the loss of the terminus signature.

The dorsal stream supplies the converse. On the two-visual-systems account (Goodale & Milner, 1992), the ventral pathway subserves conscious perceptual identification while the dorsal pathway transforms visual information into the real-time control of action. Dorsal information can scale a reaching grasp to an object’s dimensions without entering awareness: the visual-form-agnosic patient DF scales her grip accurately to a target whose size she cannot consciously report. In the present terms the dorsal output stands on the side of an origin of action and not of a perceptual terminus, so the account predicts its unconsciousness. DF’s converse pattern — ventral terminus damaged, conscious form perception lost, dorsal action-guidance spared — locates

the conscious character at the terminus signature, where the uniqueness principle places it.

The same analysis dissolves two apparent counterexamples. Motor imagery is felt because rehearsed movement arrives at the sensory termini where executed movement is felt, while the efferent command that drives real movement, being an origin, is silent (Jeannerod, 2006). The sense of agency is carried by the reafferent signals following the act, matched against prediction (Frith, Blakemore, and Wolpert, 2000), which is why it can be induced where no command issued; the readiness-potential (Libet, 1985), intentional-binding (Haggard, 2017), and illusion-of-control (Wegner, 2002) findings are the expected consequence of feeling the act, like everything else, at its terminus.

A.5 The equalisation constraint

The equalisation principle (Chapter 6) functions as a constraint on theories of consciousness, not only as confirmation of the zero-complexity constraint. The co-elimination of experience and contrast is categorical and uniform across modalities sharing no common mechanism. The person born blind has no visual phenomenology and no visual imagery, since an absent pathway entails an absent modality and not an empty one; the cochlear-implant recipient reports not silence replaced by sound but an unfamiliar dimension of experience opening.

The convergence across independent modalities with entirely different substrates confirms that the categorical character is a fact about the phenomenon, which no taxonomic convention could supply. Within a modality richness varies by degrees, while the modality itself is either present or absent, so the equalisation principle operates at that boundary.

Configuration-dependence and the argument against panpsychism. The constraint rules out theories on which experience is intrinsic to particles and not a property of configurations. If experience were intrinsic to particles, disassembling the consciousness-constituting configuration should produce gradual diminishment as each particle takes its micro-experiential contribution with it. Equalisation shows the

opposite: experience vanishes categorically when the interface conditions are no longer met; the same particles rearranged have no experience whatever. Experience belongs to the configuration; the particles carry none of it.

The panpsychist (Strawson, 2006; Goff, 2019) eliminates the boundary question by stipulating that all matter is proto-experiential, then inherits it again as the combination problem (Chalmers, 2017): what distinguishes the proto-conscious particles constituting visual experience from those constituting a rock? A sophisticated panpsychist replies that proto-consciousness is not a static shared property but combinatorial, interacting differently by configuration (Goff, 2017; Mørch, 2019).

Grant the combinatorial reading in full; two arguments then need separating. The categorical-versus-gradual argument reaches only the naive panpsychist who locates the experience equalisation removes in the particles; the combinatorial panpsychist keeps micro-experience in the particles and locates the macro-experience at the combination, so its categorical loss under disassembly is what she predicts too, so the equalisation constraint does not discriminate against her. What does reach her is the boundary question, which survives the combinatorial reading intact. What distinguishes the configurations in which proto-experiential properties combine to produce visual experience from those in which they produce a rock must cite features of the configuration — pathway, terminus, signature, interface — while proto-consciousness, present in both, contributes nothing to identifying which is which. The parallel is close: hydrogen behaves differently in H_2O than in H_2SO_4, but we need not attribute "proto-wetness" to hydrogen to explain why water is wet, since the combinatorial behaviour is fully explained by the physical properties plus the configuration. So here, the work is done by structural uniqueness, and adding proto-consciousness to the particles does no further work on the boundary question.

The panpsychist may reply that proto-consciousness answers the hard problem and not the boundary question — why there is experience at all. The chapter takes no position on that, confining its claim to the boundary question, where a property present everywhere marks nothing.

The experience was never in the particles; it was in the unique situation the particles occupied.

The identity hypothesis. The constraint favours an identity reading without establishing it: that experience and the structurally unique configuration at the interface are the same thing. In superconductivity, resistance drops to zero at a critical temperature while the system persists with rich internal structure. In the experience/contrast case, when contrast reaches zero the modality itself ceases to exist. The blind person is not a seer in a special state; she is not a seer at all. This is what identity predicts, since removing one side makes both vanish — though whether it suffices to establish identity or only to motivate it is left open.

A.6 The interface account among its neighbours

Two rival theories have the interface account as their target and no part of the hierarchy; they are met here.

Integrated Information Theory (Tononi, 2004; Tononi et al., 2016) identifies consciousness with integrated information and selects among candidate systems by maximum Φ. The two accounts address related but distinct problems. The exclusion postulate selects a system by a quantitative criterion, while the uniqueness principle identifies, by a qualitative one, what makes any system's activity experiential and not computational — and could explain why the system IIT selects has the specific character it has, which Φ alone cannot.

IIT instantiates the zero-complexity constraint and does not confirm it independently. $\Phi > 0$ requires internal differentiation, which is the presupposition the constraint identifies, so IIT's agreement displays that presupposition and supplies no evidence for it from outside (Chapter 6, §5). The point holds regardless of the major objections to IIT (Aaronson, 2014; Bayne, 2018), since it depends only on that formal requirement.

Predictive processing (Seth, 2021) holds that experience is "controlled hallucination" constituted by the brain's generative models. The accounts are compatible where predictions are pathway-specific

and prediction errors are detected contrasts, but two tensions divide them.

The first is about what is constitutive. For Seth it is the predictive model's best guess; for the interface account it is the triggered configuration at the interface, whether or not it functions as a prediction. A newborn's first visual experience, formed before experience has trained any predictive model, should on Seth's account be degraded or absent, while on the interface account it is present whenever the pathway is active against the modality's absence — a divergence the equalisation constraint resolves in the interface account's favour. The second is about quality. Identical signatures under different prediction contexts should on Seth's view differ in character and on the interface view not, which is a substantive empirical disagreement and the discriminating test of §A.7.

A.7 Falsification conditions for the interface

The uniqueness principle carries a falsification condition of its own: experience persisting when the interface conditions are absent — a motor command, homeostatic signal, or immune response constituting phenomenal experience — would fell it. The equalisation constraint adds another: experience diminishing gradually, and not categorically, as contrast is removed within a modality would fell it.

One prediction discriminates the account from its neighbours, beyond exposing it, so it is worth stating because the interface account and the predictive-processing account diverge on it. Where a dedicated pathway delivers a triggered signature at its terminus against the modality's absence, the account predicts phenomenal character even when no predictive model is in play and no global broadcast has occurred.

The clean test is direct cortical stimulation. Stimulation of the primary visual cortex produces phosphenes, experienced flashes reported as seen, with no external stimulus, no inference, and no predictive model about the world's causes (Penfield & Rasmussen, 1950; Winawer & Parvizi, 2016); dynamic stimulation traced across the

cortical surface yields form percepts without training (Beauchamp et al., 2020). On the interface account this is what a triggered signature against the modality's absence should yield, whatever the origin of the activation.

The predictive-processing theorist can reply that standing priors are engaged even by a bare flash. But a prior general enough to be always already present predicts nothing in particular, so it cannot constitute this phosphene instead of none. The global-workspace theorist has the harder case, since stimulation-evoked phosphenes can be reported under restricted attention, where the interface account predicts experience and the workspace account its absence. Recurrent-processing theory (Lamme, 2003) predicts the same dissociation of awareness from attention and report, locating feltness in local recurrence within sensory cortex; the interface account agrees on where the experience is and differs on what constitutes it: the signature against the modality's absence, which the loops of a dedicated architecture may realise without being it.

The evidence is not yet decisive, because stimulation studies vary in the attentional controls the adjudication needs. The case is nonetheless the right shape to adjudicate, where a lesser case would only embarrass; it is in that shape, and not as a settled result, that the General Conclusion's falsifiability inventory (§2) enrols it among the framework's live exposures. The equalisation constraint indicates which way it should fall: what vanishes when contrast is removed is the activated configuration at the interface, not the model or the broadcast.

Two further exposures sharpen the adjudication. Sensory substitution supplies one. Devices that deliver visual information through tactile arrays route the signal to a somatosensory terminus, so the account predicts that the resulting experience carries the somatosensory signature however visually competent the behaviour becomes — spatially informative touch, never seeing — because the signature belongs to the terminus and its contrast class. Practised users do achieve the behavioural competence at issue (Bach-y-Rita & Kercel, 2003), who read those reports as seeing with the brain and not as touch.

A substitution user who developed visual phenomenology without recruitment of the visual terminus would tell against the account.

The attentional-control gap in the stimulation literature supplies the other; the design that would close it can be specified: phosphene induction under a concurrent attentional load, with detection probed by no-report measures alongside verbal report. The interface account predicts that the signature survives the withdrawal of attention and broadcast; the workspace account predicts its elimination. Either outcome would be an adjudication, which is what enrols the design, alongside the phosphene case itself, among the framework's live exposures.

A.8 The multiple realisability objection

The objection is that identity with a specific neural configuration is chauvinistic: an octopus or an artificial system with a different substrate could not be conscious.

In reply, the uniqueness principle identifies structural conditions, not substrate conditions. Any system — biological, silicon, or otherwise — with a dedicated pathway generating a unique triggered response at an interface with absence satisfies them, because the triggered signature is defined by structural role, not physical composition. Whether "functionally equivalent" can be specified sharply enough is a question the chapter flags and does not resolve (Polger & Shapiro, 2016).

References

Aaronson, S. (2014). *Why I am not an integrated information theorist (or, the unconscious expander)*. Shtetl-Optimized, 21 May 2014. https://scottaaronson.blog/?p=1799

Bach-y-Rita, P., & Kercel, S. W. (2003). Sensory substitution and the human–machine interface. *Trends in Cognitive Sciences*, 7(12), 541–546.

Bayne, T. (2018). On the axiomatic foundations of the integrated information theory of consciousness. *Neuroscience of Consciousness*, 2018(1), niy007.

Beauchamp, M. S., Oswalt, D., Sun, P., Foster, B. L., Magnotti, J. F., Niketeghad, S., Pouratian, N., Bosking, W. H., & Yoshor, D. (2020). Dynamic stimulation of visual cortex produces form vision in sighted and blind humans. *Cell*, 181(4), 774–783.e5.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18(2), 227–247.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, 30(5–6), 481–499.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.

Chalmers, D. J. (2007). Phenomenal concepts and the explanatory gap. In T. Alter & S. Walter (Eds.), *Phenomenal concepts and phenomenal knowledge: New essays on consciousness and physicalism* (pp. 167–194). Oxford University Press.

Chalmers, D. J. (2017). The combination problem for panpsychism. In G. Brüntrup & L. Jaskolla (Eds.), *Panpsychism: Contemporary perspectives* (pp. 179–214). Oxford University Press.

Chalmers, D. J. (2018). The meta-problem of consciousness. *Journal of Consciousness Studies*, 25(9–10), 6–61.

Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200–227.

Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, 23(11–12), 11–39.

Frith, C. D., Blakemore, S.-J., & Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society B*, 355(1404), 1771–1788.

Goff, P. (2017). *Consciousness and fundamental reality*. Oxford University Press.

Goff, P. (2019). *Galileo's error: Foundations for a new science of consciousness*. Pantheon.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1), 20–25.

Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience*, 18(4), 196–207.

Jeannerod, M. (2006). *Motor cognition: What actions tell the self*. Oxford University Press.

Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, 7(1), 12–18.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, 64(4), 354–361.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, *8*(4), 529–539.

Loar, B. (1990). Phenomenal states. In J. E. Tomberlin (Ed.), *Philosophical perspectives 4: Action theory and philosophy of mind* (pp. 81–108). Ridgeview Publishing Company.

Mørch, H. H. (2019). The argument for panpsychism from experience of causation. In W. Seager (Ed.), *The Routledge handbook of panpsychism* (pp. 269–284). Routledge.

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. Oxford University Press.

Papineau, D. (2002). *Thinking about consciousness*. Oxford University Press.

Penfield, W., & Rasmussen, T. (1950). *The cerebral cortex of man: A clinical study of localization of function*. Macmillan.

Polger, T. W., & Shapiro, L. A. (2016). *The multiple realization book*. Oxford University Press.

Prinz, J. J. (2012). *The conscious brain: How attention engenders experience*. Oxford University Press.

Ramachandran, V. S., & Hirstein, W. (1998). The perception of phantom limbs: The D. O. Hebb lecture. *Brain*, 121(9), 1603–1630.

Rosenthal, D. M. (2005). *Consciousness and mind*. Clarendon Press.

Seth, A. (2021). *Being you: A new science of consciousness*. Faber & Faber.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74(11), 1–29.

Strawson, G. (2006). Realistic monism: Why physicalism entails panpsychism. *Journal of Consciousness Studies*, 13(10–11), 3–31.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5, 42.

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17(7), 450–461.

Tsuchiya, N., Wilke, M., Frässle, S., & Lamme, V. A. F. (2015). No-report paradigms: Extracting the true neural correlates of consciousness. *Trends in Cognitive Sciences*, 19(12), 757–770.

Tye, M. (1995). *Ten problems of consciousness: A representational theory of the phenomenal mind*. MIT Press.

Wegner, D. M. (2002). *The illusion of conscious will*. MIT Press.

Weiskrantz, L. (1986). *Blindsight: A case study and implications*. Clarendon Press.

Winawer, J., & Parvizi, J. (2016). Linking electrical stimulation of human primary visual cortex, size of affected cortical area, neuronal responses, and subjective experience. *Neuron*, 92(6), 1213–1219.

CHAPTER 14

Emotion as Value

Presupposes: the self-vindication mechanism (Vol. I, Ch. 1); the structural dependence of content on contrast (Vol. I, Ch. 2); the containment norm and the exhibitive closure (Vol. I, Ch. 7, §§3, 4.1); the representational explanation of the illusion (Vol. I, Ch. 8, §3); the categorical transition principle and the three-cell partition of mental features (Ch. 13, §§2.2–2.3); the reply to the universality objection (Ch. 20, §5.2), used at §A.2.

Establishes: the response-constitutive account of value; the two constitutive forms of goodness — aesthetic and empathic — with orthotic goodness as cognitive rather than modal; the identification of morality as empathic evaluation of chosen behaviour; the vacancy, containment, parity and commitment results, with the quarantine of the field; the field and the pull, which states the cost of the premise the quarantine uses; and the informed-registration anchor.

In brief. After tasting several foods, could you decide which tastes best without any emotional response? After hearing several melodies, could you rank their beauty while feeling nothing? Strip away emotion and every difference between things remains — detectable, classifiable, complete — except the difference between *better and worse*. This chapter argues that evaluative difference stands to emotion as colour stands to vision: real once constituted, response-dependent always. Its single sharpest instrument is the *vacancy result*: the clause "value obtains independently of every response" turns out to mark no contrast by any route the account can locate — a result that leaves the realist two tolled exits — and both wings of metaethics needed it, the realist to affirm and the error theorist to mourn. The surprise is where this leaves the account: not anti-realism, but realism with the one clause that says nothing removed. And the moral fact is anchored where it always was: in the affected party's own registration, not the onlooker's. One thing the account does not supply: registrations do not sum across their bearers, so the account has no first-order resource for weighing one party's agony against a million inconveniences (§6.5), and every applied verdict inherits that residue.

---

1 Introduction

After tasting several foods, could you determine which tastes best without an emotional response? After hearing several melodies, could you rank their beauty without feeling anything?

Strip away emotion and the differences between things remain — detectable, classifiable, complete — while the differences between *better and worse* vanish. The evaluative landscape goes dark.

This chapter argues that the darkness is diagnostic: evaluative difference is a distinct category of difference, and emotion is the modality whose response constitutes it.

The claim is structural, and "constitute" is chosen against "detect". Wavelengths exist without visual systems; colours do not. Colour is what obtains when a visual system responds differentially to light: real once constituted, response-dependent always. Evaluative difference stands to emotion as chromatic difference stands to vision. Physiological states, behaviours, and situations exist independently of any evaluative modality; their goodness and badness do not. The account thereby grounds the perceptual theory of emotion (Tappolet, 2016; Döring, 2007) rather than asserting it, and stands within response-dependent value realism (McDowell, 1985; Wiggins, 1987) while supplying what that tradition has lacked: a derivation, from the metaphysics of difference, of why evaluative properties are response-dependent.

The chapter's spine is the argument of §3. It asks what a value would have to be if it obtained independently of every response, and finds that the question has three answers, of which only one supplies what a value must supply, a direction (§3.5). It then asks whether the clause "value obtains independently of every response" says anything at all, and finds that it does not (§3.6), and that everything the robust realist wanted to keep survives without it (§3.7). The appendix carries the argument to the rest of the field: what realism claims as its advantage, what the error theory charges, and where each rival stands once the clause is retired.

One commitment carries that finding, and a realist will press it first, so it is stated here. Call it the anchoring requirement: a word carries no content beyond what some registered contrast supplies it, so a vocabulary that turns on itself must exit somewhere into a registered

contrast. Section 3.6 defends the requirement against the charge of verificationism, and says what its refuser keeps and what she loses.

One limit stands beside the results, because every later chapter inherits it. Registrations are proprietary to their bearers, so the account returns no fact about how one party's registration weighs against another's: there is no scale on which one agony and a million inconveniences are quantities of a common stuff (§6.5). What it supplies instead is the operation by which a trade-off is settled where anything settles it, mutual registration, converging on an outcome or, failing that, on a procedure (Chapter 15, §4.4).

Section 2 fixes the machinery: modalities, the categories their responses constitute, and, before the vocabulary of misreading is allowed to do any work, the non-teleological standard by which a response can be inaccurate (§2.2). Section 4 derives the two constitutive forms of goodness, aesthetic and empathic, and distinguishes both from orthotic assessment. Section 5 identifies morality and hands it to Chapter 16. Section 6 answers objections in ascending order of strength.

2 Categories of Difference and Their Modalities

2.1 Categories are response-constituted

Not all differences are of the same kind. The difference between red and blue is chromatic; between middle C and high C, auditory; between rough and smooth, tactile. These are not instances of a single generic "difference" but types, each individuated by the modality whose response constitutes it: chromatic differences by visual systems responding to wavelength differences, auditory by auditory systems responding to frequency differences. In each case the underlying physical variation exists independently of the modality while the categorised difference does not, since colour is a property of the visual system's response to light rather than of light itself.

This is why modalities are category-bound. Vision cannot detect auditory differences, not through an engineering limitation but because auditory differences are constituted by auditory responses. Each

modality constitutes the category it is sensitive to, and no modality can constitute another's domain.

The claim is indifferent to hardware. Vision happens to have a localised organ, emotion is distributed across neural and bodily systems, and nothing in the constitutive relationship turns on the difference. Nor is calibration variability an objection: colour perception varies with cone distributions, context, and language (Roberson et al., 2000) without ceasing to be a modality. The modality is the capacity for the constituting response; the calibration — which stimuli trigger which responses — is shaped by history, and varies.

The word "response" must be dissected before it bears weight, because ordinary usage fuses three layers the framework keeps apart. In the first, a response is a causal reaction: the thermostat to temperature, the litmus paper to acid — registration of nothing by nothing. In the second, a response is a registration: a system detects a contrast and adjusts — discrimination without anything it is like to discriminate. In the third, a response is an *oriented* registration: the contrast is not merely detected but felt as a pull toward or away: valence, the second axis of Chapter 13 (§2.3), which switches on at sentience. The response that constitutes evaluative difference is the third and only the third. "Response-constitutive" therefore names no behaviourism and no dispositional reduction: a thermostat's reactions constitute nothing evaluative, and neither would a complete causal-functional description of approach and withdrawal. Where this chapter says that emotion constitutes value, the constituting layer begins where valence begins.

One dependency completes the dissection. The account requires from Chapter 13 only the two axes, their exhaustiveness, and the existence of valence; the levels architecture frames the exposition and supplies no premise. A reader who rejects the hierarchy of levels rejects nothing this chapter uses.

2.2 Constitution without teleology

One word must be banned from the account before it does damage, since the ban is a thesis. The account will speak throughout of

misreading, of impairment, of a modality's failing its object — vocabulary that appears to measure the modality against what it is for, and functions need assigners.

Evolution, the only naturalistic candidate, assigns the wrong one. The empathic capacity was shaped for kin-biased cooperation, fires hardest for the near and the similar, goes numb at statistics, and is cheaply hijacked. A critic is entitled to observe that, graded against its selected function, the parochial empath is performing to specification and the psychopath is an alternative strategy rather than a breakdown. If the account's normative vocabulary rested on a function, it would rest on a stipulation.

The identity, not the function. It rests on an identity. A modality is individuated by its object rather than by a purpose: vision is not a tool aimed at light but the differential response to light. That is why an evolved visual system's blind spot, its narrow spectral window, and its failure in darkness are describable as limits — the object outruns the implementation, which the object grades. Colour-blindness is called a deficiency without invoking evolution's intentions or a norm of seeing. The wavelength differences are really there, the capacity to differentiate them is present in some systems and reduced in this one, so "deficient" is description rather than verdict.

Empathy stands to affected parties' registrations as vision stands to light. It is not for tracking them; it is the oriented response to them. Since the registrations are real (§6.4) and reach far beyond any implementation's native range, the implementation's parochialism, its innumeracy, and its hijackability are describable as limitation and inaccuracy on the same purely descriptive ground. Selection history explains why the instrument exists; the object fixes what counts as reading it. The telescope is explained by its makers and graded by the stars.

The consequence for the account's hardest case can be stated now and inherited later (Chapter 16, §5.2). The psychopath's condition is not empathy malfunctioning but empathy reduced or absent, less of the modality, as colour-blindness is less of vision. His verdict "nothing bad

here", delivered over a registered suffering, is false the ordinary way, since it misdescribes a fact that obtains in the sufferer. No cosmic norm convicts him and none is needed. He is inaccurate about the world; the practice responds to inaccuracy with correction and containment rather than desert (Chapters 18, 20); and the account's entire normative vocabulary — misreading, distortion, corruption — is hereby fixed as accuracy-vocabulary, answerable to objects, with no teleology anywhere in it.

One open question is owed a flag rather than a burial. The naturalised-content tradition (Millikan, 1984; Dretske, 1988) holds that misrepresentation itself cannot be defined without a function-story, in which case a residue of teleology would re-enter through the theory of content one level below the present argument. The contrastive semantics of Chapter 2 is designed to individuate content without teleosemantic apparatus, a content being a diff registered or misregistered against its object, so the account proceeds on that basis. Whether contrastive individuation fully discharges the misrepresentation problem is a question for the theory of content, recorded here as open.

3 Emotion as the Evaluative Modality

Part one: the constitutive argument (§§3.1–3.7). What the evaluative modality is, what a response-independent favouring would have to be, and why the clause that names it says nothing.

3.1 The irreducibility of evaluative difference

Consider two experiential states: the taste of a ripe strawberry and the taste of a lemon. A system with sensory modalities can register numerous differences between them: chemical composition, receptor activation, associated properties. A system with cognitive capacities can register relational differences: the strawberry is sweeter, the lemon more acidic. All of these are non-evaluative differences, differences in what the states are like. But there is a further difference, not derivable from any combination of these: which tastes *better*. A complete chemical

analysis does not deliver it; a complete neurological map does not determine it.

The irreducibility can be exhibited by a thought experiment. A being with perfect sensory and cognitive faculties but no affective capacity can distinguish every chromatic, auditory, tactile, and gustatory difference; it can classify, compare, and describe. It cannot determine which of anything is better or worse. The evaluative category is invisible to it, not for want of information (it has all the non-evaluative information there is) but for want of the modality whose response constitutes evaluative differences.

What individuates the evaluative category is orientation. Chromatic differences are unoriented: red differs from blue, but neither is "toward" or "away from". Evaluative differences are constitutively oriented: better pulls, worse pushes. This oriented character, valence, is what makes evaluative difference the basis of preference, choice, and action, which no non-evaluative category possesses.

3.2 The constitutive claim

If evaluative difference is a distinct category requiring its own constitutive modality, the modality is emotion: the affective capacity through which sentient beings generate the oriented differences between better and worse. The positive or negative valence of an emotion is the evaluative analogue of the chromatic character of a visual experience, the qualitative profile through which the modality constitutes its category. The claim is structural rather than metaphorical. Without vision there are wavelength differences but no colours; without emotion there are non-evaluative differences but no values. The modality does not accompany or facilitate evaluation; it is what makes evaluation possible.

3.3 Rival theories

The feeling theory (James, 1884; Damasio, 1994) holds that emotions are perceptions of bodily changes. The account shares its insistence on felt, qualitative states irreducible to cognition, while relocating the content: the emotion constitutes an evaluative difference,

an oriented relationship between the system and its situation, of which the bodily change is the vehicle. This is why emotions are about things — the grief is about the loss, not the tightness in the chest — which the feeling theory struggles to explain. The evaluative difference is constituted in response to a specific non-evaluative input, so the experience inherits its object from it.

The judgement theory (Nussbaum, 2001; Solomon, 1976) holds that emotions are evaluative judgements. The account grants that emotions involve cognitive content, since a fear about nothing is not a fear, while denying that the evaluative character is constituted by the cognitive content alone: a system that judges "this is a significant loss" without affective response has identified a non-evaluative fact and constituted no evaluative difference.

Recalcitrant emotions decide between the views. Nussbaum must either deny that phobias are emotions or credit the phobic with contradictory judgements. The present account gives the cleaner analysis: the cognitive system judges the stimulus harmless while the evaluative modality responds with aversion regardless, its calibration set by associative history rather than current assessment. Recalcitrant emotion is the evaluative analogue of visual illusion, two modalities delivering conflicting verdicts, and, like an illusion, it is not merely a divergent verdict but a possible misreading of the features it is directed at, in the accuracy sense fixed at §2.2 and developed at §6.4.

The perceptual theory (Tappolet, 2016; Döring, 2007) holds that emotions are perceptions of evaluative properties. The account deepens it by explaining why the analogy holds: emotions are like perceptions because both are constitutive modalities for response-dependent categories of difference. Tappolet asserts the parallel from functional profile; the present account derives it from the relationship between modality and category. The attitudinal theory (Deonna & Teroni, 2012) holds that an emotion is a felt bodily attitude taken toward a content, with the evaluative character in the attitude and none in the content. The account agrees on the location, since the orientation is the second axis and the content the first (Chapter 13, §2.3), and adds what the attitudinal

theory leaves open: that the attitude constitutes the value it is taken as tracking.

3.4 Emotional equalisation

The equalisation principle (Chapter 6, §2.2) applies to emotion as to every modality. When evaluative differences are equalised, when positive and negative valence collapse into undifferentiated numbness, evaluative content is not minimised but abolished. Numbness is not faint feeling; it is the structural boundary of feeling, as blindness is of sight. The result is the modality's own equalisation case, doing the same work here that Chapter 6 did for cognition: it fixes the lower bound below which the category does not thin but ceases. What follows in §§3.5–3.7 turns that constitutive claim against the position it most threatens.

3.5 The direction projection

The realist's value has to favour something, and this section asks what a favouring that no one registers could be made of. There are three answers only — a structure, a felt pull, or nothing — and the first does not favour, the second is the account's own, and the third is empty.

What a favouring must contain. Begin with what any response-independent evaluative fact would have to contain. Evaluative difference, §3.1 established, is constitutively oriented: better pulls toward, worse pushes away. A wrongness with no against-ness in it, a wrongness that counted against nothing, would not be a wrongness at all. So the non-naturalist's fact must carry a direction; it must favour and forbid.

The trilemma asks what that direction is. Direction is detectable by two routes that deliver different things. From outside, direction is structure: a vector, a gradient, approach and withdrawal, describable without remainder in the third person. From inside, direction is felt: the pull as undergone, the qualitative to-and-fro of valence. Either the non-naturalist's favouring is the structural direction, or it is the felt direction, or it is a third thing he must name.

The first horn is inert. A structural direction, a pattern in the space of options, obliges no one and favours nothing. It is a signpost with no push, present in the thermostat and the falling rock, and whatever authority a value has, a vector does not have it. The second horn concedes the account, since a felt direction is valence, and a favouring that exists only as undergone is response-constituted by definition.

The third horn must be produced, and the exhaustiveness that forecloses it costs nothing from Chapter 13's positive theory. Chapter 13 (§2.3) records the independence from its own side: the trilemma draws on the derived tier of the two-axes result alone, since any registered contrast either transfers across registration systems or it does not. A contrast that transfers is structure, transferability being what structure is, the shape all registration shares. A contrast that does not transfer is proprietary to its register, and a proprietary contrast carrying orientation — a toward and an away belonging to the register that undergoes it — is valence, characterised by its functional role and nothing else.

The trilemma's horns are the three cells this partition permits a favouring to occupy: transferable (the inert vector), proprietary and oriented (the felt pull), or composed of no registered contrast, which is nothing. The third is closed by the closure of Chapter 7 in its exhibitive form (§4.1): every attempted exhibition of a counterexample individuates its candidate and grids it, so the horn closes without leaning on the contested ontic thesis a realist might refuse. The exhibitive form is established, at the tier of the premise its mechanism runs through; the trilemma claims no more for it.

Two evasions close immediately. A rationalist who posits a special rational modality with a proprietary orientation of its own has described valence under a ceremonial name. One who insists on a proprietary orientation that orients no one has returned to the first horn, since an orientation with no register to pull on is a vector.

The reference relocation. A realist of sufficient sophistication will accept the compositional result and relocate his claim from characterisation to reference. The model is semantic externalism:

"water" named H_2O across centuries in which no user could characterise H_2O, because reference can outstrip description (H. Putnam, 1975). Our concept of the good, he grants, is built from response-materials through and through, yet the word may nonetheless lock onto a response-independent property, as natural-kind terms lock onto essences their users cannot articulate. On the Lewisian variant it locks on by eliteness, reference gravitating to the most natural joint in the concept's vicinity (Lewis, 1983).

The reply is that both hooks, on their working models, require what the first horn removed. "Water" reaches H_2O because H_2O does something: it causes the experiences that prompt the term's use, regulates that use, and corrects it. The reference relation rides a causal commerce between word and world, so an inert favouring leaves no trace for a term's use to be regulated by.

The magnetic hook divides by metasemantics. Where eliteness earns its pull through explanatory work, an inert joint attracts nothing, by the same concession. The view's authors hold something stronger — naturalness as primitive, reference gravitating to the world's joints because that is what reference does (Lewis, 1983; Sider, 2011) — and against the primitive version the reply narrows while a remainder stands.

The narrow reply: a magnet breaks ties among the candidate contents a term's use leaves open, and a term whose vocabulary exits into no registration supplies no candidates for ties to be broken among (§3.6). The remainder: a theorist who takes primitive reference past mediating content altogether, the word reaching the joint with nothing in between, has stepped outside the reply. The cost is a primitive normative joint written into the world's structure, the anchoring question of §3.6 relocated to the concept of that joint, with the reliability circle below unclosed. The exit is counted, with its cost, at §A.4.

The non-naturalist has one reply left, his strongest: he can refuse the demand for composition outright. Favouring, he says, is not built from contrasts. It is a primitive, an unanalysable feature that the demand for constituents misdescribes.

This relocates the question rather than escaping it. A primitive is still a determinate item — this primitive rather than that, favouring rather than disfavouring — and a determinate item is individuated by what it excludes (Chapter 2). Ask what the primitive favouring excludes, whereupon the trilemma re-forms around the exclusion. The line between the favoured and the forbidden is a structural boundary, a vector and inert; or a felt one, the pull, which concedes the account; or a third boundary the non-naturalist must again characterise, where the characterisation meets the same compositional closure it met before. Primitiveness buys exemption from analysis into parts. It buys no exemption from determinacy, and determinacy is all the argument requires.

One escape remains to be costed, because its most careful author states it explicitly and rejects the premise the reference argument deploys. Enoch (2011) denies that normative facts must do causal or explanatory work at all. The badness of agony causes nothing and is known nonetheless, our normative beliefs made reliable by a third factor: selection shaped our responses toward survival, survival is by and large good, so the shaping aimed our beliefs at the truth without the truth ever touching them.

Against this figure the efficacy constraint cannot simply be repressed. What can be pressed is the cost of declining it, which falls due in reference. A term whose use nothing regulates faces the permutation problem in its sharpest form: if neither causal commerce nor explanatory eliteness fixes which inert property “good” names, nothing distinguishes its naming goodness from its naming any of the endless inert properties that fit the same usage equally well — a difficulty pressed from within the realist literature itself (Eklund, 2017). The third factor cannot do the fixing. It secures at best that our responses would co-vary with the property if the reference were already fixed, while it operates entirely on the response side of the relation it is meant to anchor.

Two positions entered explicitly. Two positions absent from the sorting by name should be entered on it explicitly. The hyperintensional non-naturalist refines the third cell’s occupancy claim: the response-

independent favouring, she holds, differs from every response-involving property hyperintensionally rather than merely in modal profile — in what it is, in the essence idiom (Fine, 1994) — so no coextension argument touches her.

The refinement changes the grain of the claim and none of its exposure, because the trilemma never ran on modal profile. The vacancy result individuates by constitution, by what composes the favouring, and a hyperintensional difference must be a difference in something. Essence-talk that distinguishes the favouring from every structural and every valenced item leaves it composed of neither structure nor valence, which is the third cell under finer notation, with the same costs itemised and the same reliability circle unclosed. The fine grain itself is real and the framework supplies it (Chapter 2, §2.4) — for contents, where there is something to compose.

The idiom also closes over its user. An essence-attribution is a content, individuated by excluding its rivals (Chapter 2, §2.4), so the claim that the favouring differs essentially from every response-involving property is stated by drawing a contrast. The trilemma asks of that contrast what it asks of every other: structural, felt, or composed of nothing. The essence idiom refines the notation of the demand and leaves its object inside the sorting.

Humean constructivism, Street's own positive view, belongs on the list for completeness: value is conferred within each agent's practical standpoint, whose contents are contingent all the way down. That is the second cell affirmed without this account's anchor, constitution by oriented response with the answerability of §6.4 declined. The difference shows just where the anchor predicts: on the standpoint view a fully coherent standpoint indifferent to the affected party grades as well as any other, so the disagreement is localised there, at the anchor, with the rest of the partition shared.

Why the intuition is undiscriminating. Why, then, does response-independence seem so evidently true? The projection mechanism of Chapter 8 (§3.4) supplies the diagnosis, which is discriminating. A mind constituted by directional responses cannot represent a value undirected,

since a value that counted for nothing against nothing would not present as a value at all. When it represents value at all, it therefore represents structure with the pull built in: a fact that beckons, an arithmetic that obliges. The realist's intuition is what a mind of our constitution would have whether or not response-independence were true, so it cannot discriminate his view from this one — and this one explains the intuition he must take as basic.

The diagnosis has a mirror, and the accounting should show it. A mind of directional responses would also find the trilemma's partition compelling whether or not the partition were sound, so the felt evidence on the framework's side is as little probative as the realist's seeming is on his. The asymmetry between the positions is carried elsewhere: by the production challenges of §§3.5 and 3.8, whose burden lies with the producer, and by the chain of §A.4, which a challenger must reject by name. What the projection diagnosis contributes is symmetric — an explanation of each side's conviction that leaves the argument to the arguments.

The result and its condition. The direction projection is established, conditional on the response-constitutive reading of orientation (§§3.1–3.2) and the compositional premises drawn from Chapters 2 and 7, the Chapter 7 premise in its exhibitive form throughout, which is established and which the trilemma uses at that grade.

A challenger must exhibit a favouring that is neither structural nor felt, or a concept that bottoms out in no registered contrast: production challenges, with the burden on the producer. What the trilemma does not do is compel the externalist content with an inert authority. It constrains the content of his putative fact rather than his motivational psychology, so that figure is placed rather than refuted, the residual disagreement being Williams's (1981) between internal and external reasons. The costing is completed at §A.1, while the premise itself is modelled and costed at §A.5.

3.6 The vacancy result

This section turns the trilemma into a result about meaning. The realist's defining clause — value independent of every response — marks no contrast by any route the account can locate, and a clause that marks no contrast says nothing; the result is established on one stated commitment, with the self-standing family (below) left as a tolled exit.

As deployed so far, the trilemma shows that the non-naturalist cannot characterise his favouring, since every characterisation lands in a cell that defeats him. The stronger conclusion concerns the clause doing his work, "value obtains independently of every response", which fails as a thesis in a more basic way than falsity: it is vacant. For the clause to be a thesis it must be a diff, standing distinguishable from its denial in the content-level sense Chapter 2 (§2.4) fixes. That sense is wider than modal contrast, since the mathematical contents earn their standing with their modal profiles idle.

One point of order governs the audit and should be stated before it runs. A label with no content behind it is not understood and then found wanting; it is not understood. What follows does not discover, on inspection, that the clause is empty. It records why the attempt to understand it never completes — why every explanation offered for the clause hands the hearer materials that lead back into the circle or out to a response — and, with §3.5, why the attempt seems to complete all the same. The first reply to the clause is accordingly not a refutation but a request: say what would differ if it held. The routes audit is the record of the answers, and the request is put to the account's own foundation before it is put to the realist (below), so that the standard applied is one the account has met.

The width is where the natural objection lodges: if necessary truths are contents without modal contrast, the clause might be another such content, modally silent and meaningful all the same. The answer is an audit of the routes. A content secures its standing in one of three ways, with the partition of §3.5 permitting no fourth: by a modal contrast between the worlds where it holds and the worlds where it fails; by a felt contrast; or by composition from registered constituents whose

structure supplies the grain, the route by which the necessary truths survive (Chapter 2, §2.4).

The clause fails on each. By the non-naturalist's own concession the favouring is inert, so no modal contrast obtains: the two classes of world coincide in every vector, gradient, and disposition. No felt contrast obtains, since a felt difference is a response and the clause quantifies responses away. The compositional route is the one the trilemma has closed, because the clause's distinguishing constituent is the favouring itself, and a constituent composed of no registered contrast — by the partition of §3.5 and the exhibitive closure of Chapter 7 — is not a content in waiting.

It is the empty overshoot, a label with nothing behind it, so the fine grain that rescues arithmetic cannot rescue a term with no materials for grain to arrange. The response-independence clause therefore says nothing: worlds "with" the inert favouring and worlds "without" it are one class of worlds under two noises, and no route a content can travel severs the clause from its denial.

A distinction stands between saying nothing and describing the impossible; the audit must respect it, because the realist's strongest reply lives in the gap. "There is a largest prime" describes a thing that provably cannot exist, while the sentence carries content enough to be proved false, since only a contentful sentence supports a proof. Its constituents are each in good standing, and constituents in good standing compose contentful wholes even where the whole describes nothing possible; the framework itself relies on this in saving the mathematical contents (Chapter 2, §2.4).

The realist claims the same standing for his clause. "Value", "independent", and "response" are each meaningful, so the clause, he concludes, is at worst a necessary falsehood — and a necessary falsehood would serve his opponents perfectly, since the error theory needs a contentful commitment that fails. The audit owes an account of the difference between the two cases, and of which side the clause occupies.

The anchoring requirement. The account runs through a norm the grid architecture already contains. Chapter 7 (§3) orders labels, grasped diffs, and real distinctions by containment: every label attaches to a grasped diff, an inclusion that does not reverse. Read as a norm, the containment says that a label carries no content beyond what registration supplies it. Chains of inter-defined labels transmit content from word to word; they cannot create it. A vocabulary circular at its base must therefore exit somewhere into a registered contrast, or the circle passes nothing around but itself. The reading is not a gloss the account chooses to put on the containment. A label, on Chapter 7's account, is a label of a grasped diff; a label attached to no grasped diff is not a label carrying little but no label, and a vocabulary whose every explanation is another such label has nothing among its members that is a label of anything. Call this the anchoring requirement.

The foundational vocabularies satisfy it, and locating each one's exit is what gives the requirement teeth. Arithmetic is circular at its base — "two" through "successor", "successor" through "number" — and no child learns "two" from the circle. She learns it from two apples, from the registered fact that here is a thing and there is another and they are not the same thing, a pre-symbolic registration of discrete difference on which the numerical concepts are built (Dehaene, 1997). The logical vocabulary is circular, "not" through "false" and back, and negation is grasped in refusal, the spoon pushed away, before any definition arrives. The framework's own ground is anchored the same way: "difference" admits no non-contrastive definition, and what anchors it is an occurrence (Chapter 1, §5.3).

The requirement yields the instrument the question calls for, and a familiar pair exhibits it. "Heat" has two exits from its circle. It is learned through felt warmth, and independently anchored in the physics — expansion, boiling, flow, measurement — a job description a congenitally numb student could master entire from the textbook, with no feeling anywhere in it. "Hot", in its felt sense, has the one exit only.

Subtract every response from the universe and the two words meet different fates. "Heat where nothing feels" keeps the second anchor and stays contentful, and true. "Something hot while nothing anywhere feels

anything" subtracts the word's whole anchoring. Subtraction, in general, has two possible outcomes, and the term's anchors decide between them. Where an anchor independent of the one subtracted survives, content remains standing and the impossible combination is a contradiction. Where the subtracted anchor was the term's only one, the subtraction withdraws the anchoring entire, and what remains is a label with nothing behind it.

Run the audit on "value". Every response-free explanation of the term substitutes a member of its own family — "good", "what one ought to pursue", "what there is reason to want", "what counts in favour" — and each member is explained by the others, around a circle with no exit of its own.

G. E. Moore (1903) ran this test, found "good" indefinable, and read the finding as the discovery of a primitive on the model of "yellow". The reading fails at the anchor. "Yellow" escapes its circle by ostension, where the pointing lands on registered chromatic contrast. Point instead at the rescue or at the cruelty, to exhibit the response-independent value there, and everything at the end of the finger decomposes without remainder into situational fact and felt pull: the first already claimed by physics and psychology, the second a response. The open-question regress and the failed ostension report a single finding twice over — the value-vocabulary has one exit, and the exit is valence.

The residue test. The instrument now separates the clause from the mathematical cases, a separation that deserves a name. "The largest prime" inherits its content from parts anchored separately, ordering from encountered comparison and primality from encountered division, so the combination arrives fully anchored and its impossibility appears as a clash between contents both present. A contradiction requires two anchored parties; "a married bachelor" has the same structure, each word learnable from its own examples, the collision appearing only in the combination.

The response-independence clause has the other structure. Its subtraction operator takes the central term's only anchor, and what the operation leaves is a label with its anchoring withdrawn: nothing on

either side for a contradiction to form between. Call this the residue test — subtraction yields contradiction where an independent anchor survives it, and vacancy where the anchor subtracted was the only one. Necessary falsity needs a residue, and the audit finds none. The clause is vacant.

The parallel will be pressed: “married bachelor” also negates one word’s anchor with the other, since a bachelor is registered as unmarried; it is a contradiction all the same. The reply is that “bachelor” is anchored in encountered persons and “married” in encountered marriages, two anchors, so the collision is between two anchored contents; the felt sense of “hot” has warmth as its whole anchor, and “nothing feels” withdraws that anchor, so no collision occurs. The clause therefore divides by what “value” is read as. Read as valence, the only content the audit finds for the word, “value obtains independently of every response” is a contradiction of the married-bachelor kind, contentful and necessarily false, and a commitment no discourse can be shown to carry (§A.3). Read as the third-cell favouring, a value that is neither the situational facts nor the pull, it is the item the residue test finds vacant; that is the reading the realist needs, since the first hands him the account.

Call this the vacancy result. It relocates the disagreement. The account does not deny what the robust realist asserts, because no assertion is there to deny.

Two anchors on offer. A realist who grants the requirement will claim that his vocabulary is anchored, and two anchors are on offer. The first is moral experience: the circle exits, he holds, into encounters with wrongness, as the numerical circle exits into encounters with distinctness.

Audit the encounter. What is registered when anyone meets a wrongness comes to two kinds of material — the facts of the situation, the knife and the screaming, and the felt against-ness, the recoil undergone — and the against-ness is a response. The anchor on offer is the second cell of §3.5, so a vocabulary anchored there has arrived at the present account by the back stair: the constitution of §3.2 with the

label changed, and the clause's own subtraction operator severing the vocabulary from the one floor found for it.

The second anchor is a faculty: rational intuition, the direct grasp invoked where definition and ostension have both failed. Ask what its deliverance is when it arrives. Either the intuition delivers a registration, and the trilemma re-forms at once around the registered contrast, structural and inert or felt and conceded; or it delivers a seeming, and a seeming of oughtness carried with force is valence under the description §3.5 diagnoses. The most careful form of the appeal holds that seemings confer justification on their own, without any anchor beneath them (Audi, 2004; Huemer, 2005). Grant it for what a seeming is. A seeming that is neither a registration nor a pull is the third cell, which is closed; a seeming that is one of the first two lands where the trilemma places it, and the justification it confers is the justification a registration or a pull confers. No third deliverable exists for it, by the derived tier of the two axes (Chapter 13, §2.3), which is the only tier the argument draws. That the trilemma reappears intact at the level of concept-grounding, having first run at the level of the fact, is evidence of what it tracks.

Why the exit cannot be opened from the realist's side either. The failure of every explanation to reach the favouring is not a limit of the hearer's grasp but a feature of the thing. Registration is differential causal response (Chapter 13, §2.1): a contrast is registered when it makes a difference to the system that registers it. The favouring, by the realist's own concession, makes none — inert in every vector, gradient, and disposition (§3.5). So no register can be anchored in it, whatever faculty is proposed to do the anchoring; the anchor the clause needs is one that nothing could supply. This is the reliability circle of §3.5 seen from the side of concept-grounding, and §A.5 shows the two to be one cost.

The mathematical parallel will be pressed at this point. Platonists hold that numbers are causally inert and that mathematical knowledge stands regardless, so a realist about value may claim the same licence, and the challenge has been put in this very form (Clarke-Doane, 2020). The residue test is what separates the cases. Nothing registers the number two; what is registered is two apples, and the concept is

composed from that registration (Chapter 2, §2.4). The platonist's inert objects sit at the top of a construction whose every part is anchored, and the epistemology of the top is the epistemology of the parts, however far the construction climbs — the point the inaccessible cardinal makes below. The favouring sits at the top of nothing. The challenge as pressed concerns safety rather than sensitivity — whether the beliefs could easily have been false — and safety is a property of beliefs that have a content; the audit is prior to it, since it asks whether there is a content for a belief to be safe about.

The self-standing family. The remaining move refuses the requirement. Normative concepts, on the most careful statement of the position, form a self-standing family: irreducibly circular, directly grasped, standing to definition as the logical and mathematical families stand (Parfit, 2011; Scanlon, 2014). The companions carry the position, and the anchoring requirement states their cost.

The mathematical circle exits into registered distinctness, the logical circle into registered exclusion, the framework's own difference-vocabulary into occurrence. At the floor, every foundational family is anchored. What the position proposes is a family holding its companions' standing while declining the anchoring each of them has — declining it by construction, since the one anchor its vocabulary demonstrably has is the anchor its defining clause forbids. A ring of labels anchoring one another, with no anchor anywhere in the ring, does not stand where arithmetic stands. The position is not thereby refuted. What it must hold, stated in full, is a concept with no exhibition and no composition, grasped by a faculty that would have to reach an inert property without being affected by it; and a concept that answers to nothing the account can register cannot be shown empty to one who reports grasping it, only costed. §A.4 records it so, as the second of the two exits the account leaves open, and states the cost.

Granting the exemption arguendo settles nothing behind it, because the epistemic cost falls due at once. A vocabulary anchored in nothing is a vocabulary whose use nothing regulates, so the reliability circle of §3.5 stands unclosed on the far side: the shaping of normative

conviction done by selection and culture throughout, the unanchored facts contributing nothing to it (Street, 2006).

The standing of the companion survey should be marked, on the division §A.4 applies to the field. The requirement is the containment of Chapter 7 (§3) read as a norm; it holds by derivation. That every foundational family is in fact anchored is a survey, defeasible by a family it has not met. A challenger who exhibits a vocabulary in good standing whose circle exits nowhere into registration corrects the survey and leaves the requirement standing, owing at the next step an account of what anchors the exhibit.

Two objections to the requirement. Two objections press the requirement itself, and both sharpen it. The first calls the apples and the spoon a genetic fallacy: how a concept is acquired does not fix what it is about, since “fish” was learned from ostensions that included whales, and its content outran every one of them.

The observation is correct and leaves the requirement standing, because the requirement makes no claim about learning. It restates the containment of Chapter 7 (§3), which holds at a time: a label carries no content beyond what registration supplies it, whenever the audit is run. The child cases locate the exits; they do not constitute them. “Fish” illustrates the difference, since its content outran its acquisition because the term stayed anchored — encounters with the kind continued to regulate its use, and the correction that expelled the whales was itself a registration doing the regulating. A content that outruns its acquisition is still anchored. What the requirement forbids is a vocabulary with no registration anywhere in its anchoring.

The requirement will also be read as a concept empiricism, and the strongest opponent of that family is Fodor, who argued that primitive concepts cannot be learned from encountered materials, since learning by hypothesis-testing already requires the concept whose acquisition it was to explain, and concluded that they are not learned at all, being acquired by a locking that no learning performs (Fodor, 1998). The argument bears on acquisition, which the requirement does not address; what it leaves untouched is anchoring. Fodor’s own account of content

is informational: a concept means what it is nomically locked to, the property whose instances cause its tokenings (Fodor, 1987), which is a registration relation in the vocabulary of Chapter 13 (§2.1), differential causal response under another description. An innate vehicle is therefore anchored by what it locks to, as a learned one is by what taught it, and a vehicle locked to a property that makes no difference to any register is locked to no property. The self-standing theorist cannot borrow the nativism, then, because what it would supply is a vehicle, and the vehicle for "counts in favour" would still need a property that covaries with its use — the property the vacancy result finds absent.

The second objection sends mathematics back across the lines. No one has registered an inaccessible cardinal; the concept lives inside a circle of set-theoretic symbols, fixed by axioms, with no exit of its own — an unanchored vocabulary in good standing, and a precedent the realist can claim.

The reply is the compositional route the audit has carried throughout. "Inaccessible cardinal" inherits its content from parts anchored separately: "set" and "membership" from encountered collection, ordering from encountered comparison, negation from lived exclusion, assembled by operations that are themselves anchored. The transfinite is the largest-prime structure extended — full anchoring from the floor, however far the construction climbs above it. Distance from the floor costs nothing, provided the floor is there; what the requirement forbids is a circle composed of nothing outside itself. The value-family has no anchored parts for composition to assemble, since its members are one another, so the mathematical precedent belongs to the route the trilemma closed for it at §3.5.

A child who has stacked two blocks on two blocks owns every part the transfinite will ever use: collection, order, the step from a pile to a bigger pile. The tower of alephs is that afternoon's inventory iterated without new acquisitions. The value-family arrives with an empty inventory and borrows only from itself. The asymmetry is a production challenge and no assertion: the scene for number can be pointed at, the child, the blocks, the second pair set on the first, and the self-standing theorist is asked for its counterpart, a scene in which "counts in favour"

is registered as a part is registered, from which the family could be composed. The trilemma (§3.5) says why none arrives; until one does, the precedent belongs to number alone.

The companion reversed. The family's precedent claim — that normative concepts stand as mathematical concepts stand (Parfit, 2011; Scanlon, 2014) — turns against it here. Begin from a single registered difference, a mark against no mark, and a mind can generate the whole of mathematics: a second mark, a row, a plane, a solid, the operations that combine and divide them, the symbols that name a quantity or a relation among them. Every part of mathematical content is composed from differences a mind registers; the one thing mathematics requires is that the first difference occur, with no object outside the differences needed for the content (Dehaene, 1997). Mathematics is therefore no companion in inertness. It is the paradigm of content built from registration; the family that claims it as precedent inherits the construction: name the difference from which "good" is composed. A structural difference hands the property to the naturalist (§A.1); the felt difference hands it to this account; a primitive difference that is neither has to be registered by something, which is the question §A.6 puts. The realist's own semantics concedes the felt half, since on the leading realist account of how the concept is possessed, to have OUGHT is to be disposed to intend what one judges one ought to do (Wedgwood, 2007), which places the pull in the possession and leaves the property as whatever the pull answers to — the accuracy question the anchor settles (§6.4), with no third thing to answer it. What remains is to hold that goodness lies outside every difference a mind can generate, and then to owe the connection between the intuition and its object that mathematics, generated from within, never owed. That cost the realist enters on his own account as the epistemological challenge (Enoch, 2011; Street, 2006): a property that does no work in fixing the word's content, reached by a faculty whose connection to it is unexplained, with no companion left in that predicament.

The verificationism charge. The vacancy result will be called verificationism: meaning tied to a test, the untestable dismissed, the empty clause the descendant of the meaningless sentence. The charge

deserves its full form, because the resemblance is real at the surface and the differences carry the argument. The verification principle, in the form Carnap (1932/1959) gave it, held that a sentence has meaning only where some possible observation would decide it; it fell to two objections: it excluded the universal laws of science, which no finite run of observation verifies; and it failed its own test, since no observation decides the principle.

The anchoring requirement differs on both counts. It ties content to contrastive composition, with confirmation playing no part in it: a content is a diff, one of two; a vocabulary keeps its content where its terms compose from registered contrasts, whether or not any observation could decide the sentences the vocabulary forms. That is why the sentences the principle excluded are kept whole. The principle's own defenders retreated from verification to confirmation in some degree; the requirement needs no such retreat, because composition from registered parts is a requirement a vocabulary meets or fails outright, with no degree to fix. A universal law relates registered kinds; the tower of alephs is composed from encountered collection and encountered order (the inaccessible cardinal, above); the largest prime and the married bachelor are contentful because their parts are anchored even where the wholes describe the impossible. Distance from the floor is no cost to a term while the floor is there. The requirement forbids one thing only, a circle of labels composed from no registered contrast, which is the shape the residue test finds in the response-independence clause and in no other vocabulary the audit has met.

Nor does the requirement license the inference from unreportable to empty. Chapter 6 (§4.1) declines that inference where it first arises, for the pure consciousness event; the requirement declines it here: a content available in kind to registration is anchored whether or not any report of it is ever made, which is the placement Chapter 13 (§5.2) gives feltness; a content composed from anchored parts is anchored however far it lies from any test. What the requirement asks of the response-independence clause is a contrast, by any of the three routes the audit runs; the clause supplies none.

The self-application test is passed; the passing is costed at its size. The requirement is itself a contrastive content: that a label carries no content beyond what registration supplies excludes its denial; the denial, once entertained, is a diff composed from registered parts, so the requirement survives the objection that destroyed the verification principle. Passing one's own test is short of self-vindication, though, and Chapter 2 (§2.4) is why: the thesis's denial is an existential, which one contrastive instance does not kill, so the floor the vacancy result inherits is established, at the tier the Register records; a realist who resists it must resist that floor by name, at the determination dilemma's cost. What the charge leaves the realist is the position §A.4 counts as the second tolled exit, the self-standing family, held at the cost stated there. Pressed as an objection to the requirement's form, the charge lands on the older principle and leaves the requirement standing.

The grounding reformulation. The second reformulation restates the clause as a grounding claim — value is grounded in no response — and observes that grounding claims are contentful without marking modal contrasts, as Chapter 4's are. The escape fails at the relata, since a grounding claim inherits its content from the terms it relates and Chapter 4's terms are independently gridded (Chapter 7, §2). Name the property this chapter identifies, and the claim is contentful and false, to be argued where the trilemma and the dissociations wait; name the response-independent favouring, and the relatum is the term §3.5 audited, contributing nothing for the claim to govern. The reformulation borrows grounding's hyperintensional respectability for the one relatum grounding cannot supply, and either reading returns the realist to a station already costed.

The request of the point of order above is now put to the account's own foundation, since the standard would be no standard if it were applied to the realist alone. One mimicry the foundation invites makes the test sharp. Chapter 3 (§3.5) grants the terminus a determinateness that marks no property-contrast — obtaining, existential and side-less — so the realist may claim the same standing for the favouring: an obtaining beyond every contrast, licensed wherever the framework licensed its own.

The difference lies at the anchor. The terminus's obtaining is secured retrospectively, from within diversity: the occurrence of actual difference establishes that the transition from equality is actual (Chapter 5, §2.5), and the occurrence premise of Chapter 1 (§5.3) anchors the whole claim from the registration side. The framework's one contrast-free obtaining is secured. The favouring has no occurrence for anything to secure retrospectively, since every registration in its vicinity belongs already to the situational facts or to the responses, so the standing it borrows arrives without the occurrence that secured the original.

The vacancy result is established on the routes audit and the residue test. Its floor is the anchoring requirement — the containment of Chapter 7 (§3) read from the hearer's side — together with one commitment the account states rather than derives and has applied to its own terminus before applying it here: that a content is what can be exhibited or composed, and that registration, being differential causal response, cannot be had of what makes no difference. A challenger has three lodging points. She may reject the floor, at the cost the determination dilemma sets (Chapter 2, §2.4). She may hold the self-standing family at the cost §A.4 states. Or she may meet the production challenge the residue test defines: a response-free exit from the value-vocabulary, an anchored remainder for "value" that survives the subtraction as the physics of heat survives the subtraction of warmth. Nothing of that shape is on record, and the open-question regress is the record of the search.

3.7 The containment result

The vacancy result supports a second, which reverses the account's apparent position on the field. Robust realism about value bundles four commitments: evaluative claims are truth-apt; some are true; their truth is independent of the observer's take; and their truth is independent of every mind whatsoever.

The account secures the first three outright. Evaluative claims state response-dependent facts and are true when the relevant properties obtain (§6.2). The moral case is anchored in the affected party's own registration, which obtains whatever any onlooker feels or fails to feel

(§6.4): observer-independence without mind-independence. The realist's strongest intuition — that her agony is bad whether or not any judge says so — is therefore true and is the account's: the badness is independent of the observer because it is constituted by the sufferer, so the intuition is evidence for the anchor and none for the clause, which adds "every" and deletes her. And error is real, since a response's content component can misregister its object (§6.4 on fittingness).

The fourth commitment is the response-independence clause, and §3.6 shows it vacant. The containment result follows: response-constitution contains every contentful commitment of robust realism, and what it declines is only the residue that says nothing. The account is therefore misdescribed as an anti-realism making concessions to realist intuitions. It is realism with the vacant clause removed — everything the realist can state, kept; the part he cannot state, returned to him.

Where the argument stands: the modality has been identified and its normative vocabulary fixed as accuracy-vocabulary (§2), the trilemma has sorted every possible favouring into three cells (§3.5), the vacancy result has emptied the third (§3.6), and the containment result has shown response-constitution keeping everything the realist can state (§3.7). Part two audits what the rest of the field can carry.

Part two, in the appendix: what the field will and will not carry (§§A.1–A.5). The naturalist absorbed, the realist's claimed advantage audited, the error theory dissolved, the field quarantined, and the premise beneath it all modelled and costed.

4 Three Forms of Goodness

With the field closed, the remaining work is internal: what forms the constitution takes. The taxonomy that follows divides goodness by the object of the constituting response, and everything the applied chapters use — the empathic identification of morality (Chapter 16), the two registers of criticism, the political derivation (Chapters 21–22) — is drawn from it.

4.1 Aesthetic goodness

Aesthetic goodness is the evaluative modality responding to the outputs of other sensory or representational modalities. When we perceive a sunset and find it beautiful, the visual modality constitutes chromatic and spatial differences; the evaluative modality responds with an oriented attraction. Beauty is an evaluative difference constituted by emotion in response to what vision constitutes. This is why aesthetic judgement is irreducibly emotional: one can describe a painting's composition in exhaustive technical detail without capturing its beauty — the description catalogues non-evaluative differences; beauty is the evaluative difference constituted in response to them.

4.2 Empathic goodness

Empathic goodness is the evaluative modality responding to the emotional states of others, or, in the self-empathic case, to one's own. The step is sharp. In aesthetic goodness, emotion responds to what is not emotion; in empathic goodness, emotion responds to emotion. This is a categorically different operation; it is why empathy is irreducible to cognitive perspective-taking.

One can know, in full cognitive detail, what another person is feeling without responding emotionally to that knowledge. The perspective-taker has orthotic access to the other's state, since she can check the description against the evidence, without empathic access. Empathy requires the evaluative modality to be engaged, and this engagement is what constitutes care, compassion, and moral concern. Like aesthetic response, it may be direct — tears, a smile — or indirect, the thought of what another may be feeling.

4.3 Orthotic goodness

Orthotic goodness, from Greek orthos, straight or correct, is goodness as conformity to a standard. A correctly spelled word, a valid argument, a compliant action: each matches a pre-established standard, and the matching can be checked without any affective response, since a spell-checker assesses orthotically without emotion.

Orthotic goodness confirms the modality argument by contrast. If all forms of goodness required the evaluative modality, the constitutive claim would be trivial. It is substantive because assessment, cognitive comparison against a standard, and evaluation, the constitution of oriented difference through response, are different operations. An emotional response may follow an orthotic judgement — a misspelt word may look unpleasant — and the response that follows is aesthetic: orthotic assessment provides the object, while aesthetic or empathic response provides the value.

4.4 Boundary cases

Two boundary cases fix the taxonomy's joints, and both are used by later chapters. *Disgust at injustice* seems empathic but is triggered by a conceptual recognition. The taxonomy identifies it as empathic evaluation mediated by cognitive content: the cognitive system identifies the injustice; the evaluative modality responds with oriented repulsion to the suffering the injustice produces. The mediation does not make the evaluation orthotic, because the felt wrongness is constituted by the response, not by the recognition of rule-violation. *Aesthetic appreciation of moral conduct*: a firefighter's rescue is morally admirable (empathic evaluation of the consequences for the stranger) and may also be aesthetically admirable (response to the act's form and grace). The two evaluations co-exist without collapsing, because they take different objects: remove the empathic component (the building was empty) and the moral evaluation lapses while the aesthetic one may survive. Chapter 16 (§2) turns this object-difference against the aesthete of cruelty.

5 Morality as Empathic Evaluation

The taxonomy generates a definition, developed and defended in Chapter 16: morality is empathic evaluation of behaviour that stems from sufficiently complex choice. Three conditions are jointly necessary. The behaviour must be chosen, since an asteroid that destroys a city is not morally evaluated. The choice must be sufficiently complex, since a dog that bites on instinct behaves badly but not

immorally; the complexity has its address on the output axis of Chapter 13 (§2.5), where the act must run through the alternatives and not by the instinctive route, in a field in which the affected party is registered as a subject. And the evaluation must be empathic — the modality responding to the emotional consequences of the conduct for others, rather than to its form (aesthetic) or its conformity (orthotic).

The response that makes such conduct good or bad is, in the first instance, the affected party's own: her registration of what is done to her, with the evaluator's empathy the modality by which that registration is read from outside, accurately or not (§6.4). The distinction between moral evaluation, which is empathic, and moral rule-following, which is orthotic assessment, is the chapter's most consequential result for ethics, and its systematic conflation by the major traditions is the subject of Chapter 16.

6 Objections and Replies

6.1 The rationalist objection

Objection. Reason can access evaluative difference without emotion: we can know that torturing innocents is wrong by rational reflection alone: applying principles, testing universalisability, calculating consequences.

Reply. The objection conflates two operations. Reason can determine that torture violates a principle, fails a test, or produces suboptimal consequences; these are orthotic assessments, and real cognitive achievements. What they do not do is constitute an evaluative difference: a system that checks behaviour against the categorical imperative without any response to the suffering involved registers conformity or non-conformity, and generates no oriented pull of moral concern. So Kantian principles are orthotic instruments, tools derived from prior emotional responses to suffering and flourishing, so their moral authority rests on the modality rather than on reason alone. The rationalist's report that he feels no pull while he judges is granted; §A.6 places the cold verdict as an anchored representation deployed while the

register rests, so the report supplies no third source for the direction the verdict carries.

Schroeder (2007) grounds reasons in desires, and the account supplies what that reduction leaves open: reason-sensitivity is grounded in the modality's oriented response, since a consideration that engages no evaluative response is motivationally inert, and a reason that moves no one is not functioning as a reason. The exchange ends, as §3.5 recorded, at the internal/external reasons divide (Williams, 1981; M. Smith, 1994). The trilemma constrains the content of the rationalist's putative response-independent fact rather than the psychology of the externalist who accepts an inert authority, and that figure is placed, at the cost itemised there.

6.2 The projectivism objection

Objection. Mackie (1977) and Blackburn (1993) agree that evaluative judgements are grounded in emotional responses while denying that the responses constitute real properties: values are projections of attitudes onto a world that contains none. If quasi-realism succeeds, the response-constitutive account adds only an unnecessary ontology.

Reply. The projectivist and the account agree on the psychology and divide on the ontology, and the division is settled by the trilemma rather than by analogy. The colour case supplies the picture: colours are constituted by visual systems responding to wavelengths (Hardin, 1988) and are not thereby unreal. The projectivist misdescribes this relationship as projection onto a world of bare wavelengths, whereas the visual system constitutes colour rather than projecting it. But the value argument depends on no verdict about colour, which is itself contested; the warrant is internal to value (§3.5).

What projectivism cannot do, and the account can, is distinguish a response that constitutes a real property, the empathic response to suffering, from one that merely occurs, habitual disgust at the unfamiliar. The distinction is drawn by fittingness (§6.4), and D'Arms and Jacobson (2000) articulate its shape: evaluative properties are

attributable to their objects rather than projected when the constituting response is fitting. Street's (2006) Darwinian debunking argument confirms the account's position from the flank it targets, since her dilemma bites theories on which attitudes track attitude-independent truths, and nothing here is tracked. The evolutionary shaping of the modality explains its calibration without threatening the reality of what it constitutes, as the evolutionary origin of colour vision explains which wavelengths are discriminated without making colours illusions.

6.3 The evolutionary and cultural objections

Objection. Emotions evolved for survival, not for evaluative truth; and emotional responses vary across cultures, so the account entails that values vary with them.

Reply. The evolutionary point answers a question about origin rather than constitution. Selection explains why organisms possess the modality, just as it explains why they possess colour vision, and shows neither category illusory (§6.2). The teleological version of the worry is answered at §2.2: the modality is individuated by its object rather than by its selection history; the object grades the implementation.

On variation: yes, evaluative differences obtain where responses obtain, and the same dish is really delightful in one context and really disgusting in another. Two things prevent collapse into relativism. The constitutive role is invariant, with only the calibration varying, shaped by development and culture. The moral case is anchored (§6.4): the fact that settles a moral evaluation is the affected party's registration, which cultural calibration on the evaluator's side can read accurately or miss, without creating it. Cross-cultural convergence on the core cases — cruelty to children, unprovoked violence (Schwartz, 2012) — is what the anchoring predicts. Where calibrations conflict, the mechanism of adjudication is deep persuasion (Chapter 21) rather than a further fact.

6.4 The fittingness objection

Objection. The account leans repeatedly on fittingness: a response constitutes a real evaluative property only when it is fitting (§6.2). But fittingness is a normative notion, and the wrong-kind-of-reasons

problem is that it cannot be grounded in the responses themselves without circularity. To call a response fitting because the modality delivers it is to say nothing; to call it fitting by a further standard is to smuggle in the stance-independence the account has rejected, or to start a regress of responses.

The objection bites hardest at the coherent outlier: a being whose empathic modality is intact but runs in reverse, so another's suffering attracts it. On the account, this being constitutes, for itself, the property that cruelty is good. If its response cannot be pronounced unfitting, the outlier cannot be pronounced mistaken, and the chapter's ethics helps itself to a standing it never earned.

Reply. The account cannot supply a stance-independent fittingness and should not pretend to. It does not need one, because the work fittingness was wanted for divides into two cases, and a single relocation settles the first outright and shrinks the second.

The relocation. Moral evaluation takes as its object the emotional consequences of conduct for others (§5). For moral value, then, the response that constitutes the evaluative difference is not in the first instance the evaluator's; it is the affected party's. Her suffering is her own modality registering what is done to her as bad: a response-dependent property that obtains independently of any onlooker. It is observer-independent without being mind-independent — remove her mind and the disvalue is gone, remove every onlooker and it remains.

The evaluator's empathy does not constitute this disvalue. It is the modality by which another registers it (§2.2), and, like any modality trained on a real feature, it can register accurately or fail to. This settles the cases the later chapters prosecute. The slaveholder's comfort is not a rival constitution of value standing level with the abolitionist's horror. It is a response that survives only by not taking up what is there, her suffering, already registered by her own modality; it does not become accurate because his society shares it, any more than a wall is red because a roomful of people view it under a red lamp. The verdict is not condemnation from our side but error: a response failing to register a real, response-dependent feature of its object.

The generalisation converts fittingness from an unexplained normative primitive into a structural notion the two axes already contain (Chapter 13, §2.3). Every evaluative response decomposes into two components. The content component is the registration of the object's features, assessable for accuracy without invoking any value. The orientation component is the valence coupled to it, which constitutes the value and, representing nothing, cannot be assessed for accuracy.

A response is fitting when its orientation is coupled to an accurate registration, and unfitting when the registration misrepresents: the phobic's fear registers a threat the aircraft does not pose; grief over a false report registers a loss that has not occurred; the slaveholder's comfort survives only by failing to register what is there. Nothing normative has been smuggled in, since accuracy is a representational notion and the account has been entitled to representation since Chapter 7.

The wrong-kind-of-reasons problem receives its answer from the same decomposition. The demon who threatens harm unless the saucer of mud is admired (Rabinowicz and Rønnow-Rasmussen, 2004) supplies a reason to have the admiration while bearing not at all on what the admiration's content gets right. Reasons of the right kind bear on the accuracy of the content component; reasons of the wrong kind bear on the value of hosting the response. They belong to different axes rather than to different weights on one scale.

The floor of the regress. The regress has a floor. What corrects the sufferer's own response? An evaluator's empathy is a view taken on another's state, and a view can misrepresent; adaptive, distorted, and manipulated responses, on the evaluator's side and the sufferer's alike, are failures of just this kind, exposed by convergence and by intra-agent consistency (Chapter 15).

What is not a view taken on a further object is the sufferer's registration of her own occurrent suffering. There is no gap between the pain and its being felt as bad-for-her across which a misreading could open, since the registration is the undergoing of the state. Distortion can reach what she believes about her condition, her options, or her worth

— views, correctable as views — without reaching the badness of the suffering as it occurs. That occurrent badness is the anchor: the one response in the chain not itself read off anything further.

The clinical dissociations. The clinical dissociations raised against the floor confirm its placement. Pain asymbolia — nociception intact, suffering absent — is the two-axes architecture in clinical form: valence couples to content and can decouple, the floor is laid under the affective component alone, and where there is no suffering the anchor makes no claim.

The masochist is a composite of Chapter 15's kind. The pain's negative primary component is real and stays negative, since it is what the practice requires, while the total is carried positive by added value. Sought pain is badness outweighed, never pain registered as good.

Four further positions. Each presses on the floor, and one fork meets them all. Alexithymia (Taylor, Bagby, and Parker, 1997) and the unreliability of introspection (Schwitzgebel, 2008) drive a wedge between the felt badness in its occurrence and the subject's fallible reports of it, which is where the floor requires one, since it claimed authority for no judgement. The higher-order theorist (Rosenthal, 2005), taken as a rival account of feltness, is met at Chapter 13 (§3); taken arguendo he delivers less than the objection needs, since on his own terms the badness-for-her occurs, with the same reality, whether or not a first-order target exists. The evaluativist about pain (Bain, 2013), for whom unpleasantness is the pain's representing a bodily condition as bad, meets a fork on the represented badness: response-independent badness of the condition relocates the vacant clause (§3.6) into the body, while badness-for-the-creature is the valence axis under a representational description, so the position divides into a content component the floor never claimed and a badness component that is vacant or is the undergoing — absorbed on the pattern of §A.1. Klein's (2015) imperativism keeps the oriented component under another heading, since a command felt with no pull would be the inert entry the two-axes architecture excludes; and Corns's (2014) disunity would matter only if the anchor quantified over a kind, where it quantifies over episodes, each carrying its own valence and grounding its own fact.

What would breach the floor in every case is suffering that occurs while nothing is undergone, the item Chapter 13 (§2.3) challenges the critic to produce, and no case supplies it.

The real residue is the outlier who does not misread: who registers her suffering with full clarity and is drawn to it. Here the relocation does its second work. His attraction is his modality's response to her state, while the moral fact about her treatment is constituted by her response rather than his. His response constitutes, at most, that her suffering is good for him — a fact about his satisfaction, which says nothing about the consequences of his conduct for others; that is the whole of what moral evaluation concerns.

Asked the moral question — good or bad, reckoned by its effect on her? — the answer is settled by her; it is bad. The outlier who answers "good" has redescribed his self-interest as a verdict, or changed the subject.

What the account still cannot do is compel him. That his conduct is bad is a verdict he may grant and flout, and against his motivation, rather than his verdict, the account offers what the book offers everywhere: opposition, and where possible deep persuasion. This is a small residue, owned in the open. His verdict can be called mistaken and his self-description confused; only his will lies beyond reach, where, on every theory ever proposed, it always lay.

Both results on which this reply depends — the directedness of evaluative responses and the anchoring of moral disvalue in the affected party — follow from the two-axes architecture rather than resting on the dialectic alone. Valence is never free-standing: it is the orientation of a content, and contents have accuracy conditions (Chapter 2), so an oriented response inherits the answerability of the content it rides. Where the evaluation is empathic, its individuating object is the affected party's felt condition (§4.2), so answerability to that party is the constitutive accuracy condition of the response's own content rather than a courtesy appended to it.

The account's other wrong-kind instrument is complementary rather than duplicate. The route partition of Chapter 15 (§3.5) sends the

demon's reasons for admiring the saucer of mud down the consequential channel while admiration itself is constituted on the primary route. The present reply fixes which axis a fittingness verdict grades; the route reply fixes which channel a reason arrives by; and a wrong-kind case can fail on either.

6.5 The silent-harm objection

Objection. The account's anchor is the affected party's felt registration (§6.4); it is the account's answer to relativism. But there are wrongs, or apparent wrongs, in which registration is absent, and the anchor falls silent where a theory is most needed. Four cases fix the shape.

The undiscovered betrayal: a friend despises you, mocks you to others, reads what was private, for forty years, undetected; you die content, felt registration zero, yet the verdict of wrong is among the firmest the practice returns. Posthumous wrongs: the promise to the dying broken after the death, with the party registering nothing. The blissful violation: an agent covertly drugged into stable contentment while her commitments and relationships are taken over, her registration positive, so the anchor appears to bless the violation. And aggregation: one party's agony against a million inconveniences, where registrations exist but each is proprietary to its bearer, and no scale sums them. Fittingness (§6.4) polices the evaluator's reading of a registration that exists; none of these is a misreading. The anchor looks like a fair-weather instrument.

Reply. The resources are already in service, so the reply's work is to state canonically what the applications use piecemeal. The tests for adaptive preference (Chapter 15, §6.6) and manufactured added value (Chapter 15, §7.2), the fit criterion for purpose (Chapter 19, §6.3), and the manipulation audit (Chapter 21, §7.6) share one form: a response, or its object, is assessed against the party's own registration corrected only for accuracy and completeness about its object.

The informed-registration anchor. Stated as the account's second anchor: the moral fact about a party's treatment is fixed by her own

registration of it — actual where actual registration is accurate and complete, counterfactual where it is not, the counterfactual being her modality's return under full and accurate contact with what was done.

Its discipline is threefold, because every counterfactual is a step toward the ideal-observer standpoint the account forbids. One residue remains. "Accurate and complete registration of the object" is not a determinate specification: how much acquaintance counts as complete, and against whose description of the object, are questions the anchor does not settle from inside. So where a counterfactual registration decides, the anchor fixes where the moral fact lies, and fixes its content only as far as the completeness it invokes is itself determinate — settled in the easy cases, open in the hard ones. The correction is minimal: accuracy and completeness about the object of the treatment, with no further idealisation of sensibility or information. The modality is hers, so the counterfactual is indexed to the party's own evaluative constitution and no consensus, ideal responder, or theorist enters under the subjunctive; where her corrected modality does sustain a response, the verdict stands however unusual, as with the surviving grass-counter (Chapter 19, §6.3). And the counterfactual is diagnostic only: where actual registration is accurate and complete, it is the fact, and no would outranks an is that misses nothing.

The betrayed party first. What was done to her was the manufacture of a systematic misregistration: for forty years her trust and warmth were keyed to a friendship that was not there, her evaluative life run on fabricated inputs, so the occurrent anchor is silent because the silence was engineered, and an engineered silence is part of the fact — the manipulation analysis of Chapter 21 (§7.6) applied to a whole relationship, a stasis sustained by concealed registrations whose exposure would dissolve it. Her informed registration is what the concealment existed to prevent, and it returns the verdict of betrayal with full force: the wrong is the manufactured gap between her felt world and its objects, its magnitude the registration her own modality returns under accuracy.

One residue is owned. Where the illusion is perfect and permanent, the account locates the wrong in falsified registration, with no entry

under felt pain. A critic whose intuition demands felt harm is invited to press the intuition, which under pressure standardly resolves into the account's terms: asked why the undiscovered betrayal is bad, the practice answers that the friendship was not real.

The posthumous case divides, which is the anchor working. Where the party registered, while living, what would be done — she cared, occurrently and accurately, about her ashes or her word — the promise engaged registrations that occurred, so the breach falsifies them retrospectively: the betrayal structure, with the gap opened after registering has closed. Where living registrants remain, the ones who loved her and the participants whose registrations depend on the reliability of promising, their registrations carry the remainder.

Where neither obtains, the account returns a verdict on the promise-breaker and no patient-wrong, because there is no patient. Some of what the practice files under wronging the dead is, on this account, the projection of §3.5 in mourning dress, the empathic modality firing at a person-shaped absence: an intuition predicted and declined.

The blissful violation costs nothing once the discipline is applied, because it is the adaptive-preference structure with the adaptation administered by an agent. Run the audit: would her contentment survive accurate and complete registration of what was done — commitments overwritten, relationships selected, consent bypassed? It would not. The contentment lives on concealment as the adaptive preference lives on foreclosed alternatives, so the informed anchor returns the violation's badness in her own voice, the verdict recovered from her rather than imposed by a theorist. The case that looked like the anchor endorsing the violation shows why the anchor was always two-part, since an engineered registration is what the accuracy correction exists to see through.

Aggregation differs in kind, and the account answers with a location. Registrations are proprietary to their bearers, by the partition of §3.5 again, so there is no inter-register fact of the sum: no scale on which one party's agony and a million headaches are quantities of a common stuff. The refusal to sum has a classic statement in Taurek

(1977), reached there from the standpoint of each person's own loss, which no sum is a loss to; here it follows from the proprietary character of registration. The developed procedural answer is Scanlon's (1998): a principle is tested against each person's complaint, never against the sum, which is a rule for deciding where the registrations give no fact to read off, credited as such at Chapter 16 (§4.2). What exists is the several registrations, each real and anchored, together with the question of how conduct should answer to all of them at once — the constitutive question of the political layer (Chapter 15, §4.4; Chapters 21–22).

The utilitarian pressing the objection should be pressed back on his instrument. The sum is a construction, an orthotic device built to make registrations commensurable for decision, and mistaking the device for a discovered fact is the great moral conflation in quantitative dress (Chapter 16).

A closing note fixes the anchor's scope. It is indifferent to substrate: drawn at the level of registration, it inherits the ethereal mind argument (Chapter 18) and the standing criterion's phrasing, the capacity to feel counting wherever it is found (Chapter 22, §4.1). A simulated party's suffering, being undergone, is suffering, so a global simulation subtracts nothing from the verdicts above; where value lives if the world is simulated has one answer, where the feelers are.

7 Conclusion

If the argument of this chapter is correct, value is not discovered in the world like a fossil in rock. It is constituted by the modality's engagement with the world: as real as colour, as dependent on the responder, and as indispensable for navigating a world that would otherwise be evaluatively dark.

The chapter's results are tabulated below with their grades and rowed in the Register. Its defence against the field is the Appendix, where every stated rival is met, the two tolled exits are counted, and the one contested premise modelled and costed with the burden on the refuser.

The eight results that meet the field, mapped, with the section that states the cost of the premise each uses:

Result	**§**	**Grade**	**In one line**
Direction projection (trilemma)	3.5	established, conditional	A favouring is structural, felt, or nothing; only felt directs
Vacancy	3.6	established	“Independent of every response” says nothing
Containment	3.7	constructed	Realism kept entire, minus the clause that says nothing
Absorption	A.1	constructed	The naturalist holds this account under another name
Parity	A.2	established	Nothing anywhere binds the unmoved; absence here discriminates nothing
Commitment	A.3	established	The error theory’s alleged commitment was never available
Quarantine	A.4	established (negative half)	No contentful rival occupies the field; the survey is defeasible
The field and the pull	A.5	established, with a noted dependency	The premise the externalist refuses, modelled and costed; burden on the refuser

The support remains a braid rather than a chain. The trilemma is a conceptual elimination. The projection diagnosis is a debunking line that stands if the trilemma is resisted. The clinical dissociations (§6.4) are an empirical line: where valence is selectively removed, value vanishes with the response, which constitution predicts and tracking must treat as coincidence; where the register is silenced on the judge's side, the verdict survives on its anchoring while its motive fails and its edge cases drift (§A.6), which the same constitution predicts. The vacancy result is a semantic line that survives the contesting of every empirical strand. The informed-registration anchor (§6.5) extends the occurrent anchor to the silent harms under a stated discipline, its residues stated in the open, with the dumbfounding data sorted by the same instrument (§A.9).

Two registers of criticism follow, and both are used throughout what follows. The empathic verdict that conduct is wrong is anchored, in the moral case, in the affected party's own response, a real feature an evaluator's empathy can register accurately or miss. The orthotic judgement that it is incorrect is relative to a standard. No third, stance-independent register is available to any party. This distinction is what remains of the sentimentalist tradition's insight once it is given a metaphysics: where Hume (1751/1998), Blackburn (1993), and Prinz (2007) ground moral judgement in response, the present account grounds the response's category, its object, and its accuracy conditions.

Morality as empathic evaluation (constructed). Morality is empathic evaluation of behaviour stemming from sufficiently complex choice: an identification stated here and prosecuted in Chapter 16. It raises a prior question. Evaluation involves multiple responses firing simultaneously — to the object, its associations, its origins, its consequences. How are competing responses resolved, and what transforms a web of associations into a hierarchy of value from which choices flow? Chapter 15 develops the answer, which will prove essential for understanding how the major ethical traditions went wrong.

Grades and Dependencies

1. **Grades.** Nine results. The response-constitutive account of value is *constructed*, on category-individuation by modality (§2.1) and orientation as the evaluative mark (§3.1); it falls with the trilemma's second cell, since the eliminativist must exhibit suffering nothing undergoes. The direction projection, the trilemma, is *established*, on §§3.1–3.2 and the compositional premises of Chapters 2 and 7 in exhibitive form; what would break it is a favouring neither structural nor felt — a production challenge. The vacancy result is *established*, on the routes audit and the residue test, with the anchoring requirement from Chapter 7 (§3); its compositional route, the residue test, and the commitment result fall if the anchoring requirement is refused, at the cost the Register's contested-premises list states, while its modal and felt routes stand on the determination dilemma (Ch. 2, §2.4). The containment result is *constructed*, on §3.6 plus the four-commitment audit of robust realism, and falls with the vacancy result. The absorption result is *constructed*, on the naturalist fork (§A.1); what would refute it is a naturalist fact that is neither an oriented response nor an inert residue. The parity result is *established*, on one descriptive premise — that assent-compulsion is exhibited by nothing; what would refute it is an agent turned by a requirement while every response channel is flat. The commitment result is *established*, on §3.6 in its strong reading and the Chapter 7 biconditional; it requires strong contentlessness and falls on a weakened vacancy reading. The quarantine of the field is *established* as a partition and defeasible as a survey, on the Chapter 1 mechanism, the exhibitive closure of Chapter 7, and the partition of §3.5; a rival occupying no cell corrects the survey and leaves the partition standing. Morality as empathic evaluation is *constructed*, on the goodness taxonomy (§4), and is prosecuted at Chapter 16; it falls with the taxonomy or the moral identification. The aggregation limit (§6.5), that registrations do not sum, is a limit the account states, carried in the Register as the trade-off residue (Ch. 15, §4.4) beside the contested premises; no result falls with it, and every applied verdict on a conflict of registrations is read under it.

2. **Chains and exits.** The positive results are constructed; three negative results — direction projection, parity, commitment — are established, each on a chain with no constructed link; the vacancy result is established on the same path together with one commitment §3.6 states rather than derives

(content is what can be exhibited or composed; registration is differential causal response), applied to the account's own terminus before the realist's. The quarantine divides: the partition of possible locations is established by logical form, while the survey of actual occupants is defeasible by a position it has not met. Two exits stand outside the field at stated cost, the primitive-structure magnetist (§3.5) and the self-standing-circle intuitionist (§3.6). One premise is contested by name — that favouring is grip — modelled and costed at §A.5 and carried in the Premise Register with what falls and what stands if it is refused. The cold verdict (§A.6) is placed as a defeated objection and carries no row: it rests on the anchoring requirement (§3.6), the diagnosis of §3.5, and the model of §A.5, establishing no thesis. The anchoring requirement is stated with the results (§1) and defended against the verificationism charge at §3.6, with what a refuser keeps and loses listed at both places. The self-standing family's exit is costed at the epistemological challenge the realist enters on his own account, carried without the mathematical companion (§3.6).

3. **Instruments received and used.** The two axes and their exhaustiveness, at the derived tier, come from Chapter 13 (§2.3) and are used at the trilemma (§3.5) and the intuition audit (§3.6). The existence of valence as the orientation of a content comes from Chapter 13 (§2.3) and is used at §§3.1–3.2 and the anchor (§6.4). The exhibitive closure comes from Chapter 7 (§4.1) and is used at the third-cell closure (§3.5) and the vacancy audit (§3.6). The containment of labels by grasped diffs comes from Chapter 7 (§3) and is used at the anchoring requirement (§3.6). The contrastive individuation of content at fine grain comes from Chapter 2 (§2.4) and is used at the routes audit (§3.6) and the hyperintensional reply (§3.5). The projection mechanism comes from Chapter 8 (§3.4) and is used at the diagnosis of the realist's seeming (§3.5) and at the cold verdict (§A.6). The reply to the universality objection, run in reverse, comes from Chapter 20 (§5.2) and is used at the parity result (§A.2). The as-if loop comes from Damasio (1994) and is used at the tense of the cold verdict (§A.6). The possession account of OUGHT comes from Wedgwood (2007) and is used at the companion's reversal (§3.6), where the realist's own semantics places the pull in the possession of the concept. The Premise Register carries the chain, with the trade-off limit among the residues it records.

Appendix: The Quarantine of the Field

This appendix holds the account's defence against the field: the results that meet the stated rivals to response-constitution, each rowed in the Register at its own grade, and none of them used by the chapter's positive account except through the results named at §§3.5–3.7 and §6.4. Its sections are numbered A.1–A.9 and cited as §A.n. The table maps the field — each position the account meets, where it is met, the verdict, and what holding it costs — and the sections that follow are the arguments, in the order the chapter's exposition ran them: three results (§§A.1–A.3) that the quarantine (§A.4) assembles, two sections (§§A.5–A.6) that state the cost of the premise the quarantine uses, and three replies (§§A.7–A.9) to objections the quarantine sends to the account's own construction.

Position	**Met at**	**Verdict**	**What it costs to hold**
Non-naturalism (Parfit, Scanlon, Shafer-Landau): a favouring that is neither structure nor felt pull	§§3.5–3.6; §A.2	Placed	Clause vacant, reference unfixed, reliability circle unclosed, authority binding no one
The self-standing family: the intuitionist's circle of normative concepts	§3.6; §A.4	Tolled exit	A concept with no exhibition and no composition; a faculty that would reach an inert property; a use nothing regulates; the circle unclosed
Primitive-structure magnetism: reference without mediating content	§3.5; §A.4	Tolled exit	A primitive normative joint; the reliability circle unclosed

Position	**Met at**	**Verdict**	**What it costs to hold**
Constitutivism (Korsgaard): value conferred by the reflective structure of agency	§A.4	Placed	Endorsement that engages is valence, with the practical identity as its content; endorsement that only registers is the first cell, and shmagency is the demand for the pull it lacks
Naturalism (Cornell realism; "badness just is suffering")	§A.1	Absorbed	Restates the account, or posits an inert further property the trilemma has already costed
Quasi-realism (Blackburn, Gibbard)	§6.2; §A.4	Converged	The account's endpoint reached by a longer road, at the cost of one idle step
Creeping minimalism	§A.8	No purchase	The account's distinctness is ontological, so the collapse of the realist/anti-realist vocabulary leaves it untouched
Error theory and fictionalism (Mackie)	§A.3	Dissolved	The commitment the error requires is impossible; fictionalism inherits the dissolution
Eliminativism about valence	Ch. 13, §2.3; §A.7	Silent	Must exhibit suffering that occurs while nothing is undergone; a being without the responses issues no objection

Position	**Met at**	**Verdict**	**What it costs to hold**
Conceptual-role realism (Eklund)	§A.4	Placed	The role's only working occupant is the felt pull: the vocabulary kept, the realism gone in the naming
Argued quietism (Dworkin)	§3.7; §A.4	Contained	Receives the face value it wants — truth-aptness, truth, observer-independence — minus the clause that says nothing, which it cannot mourn
Strict quietism: refusal to engage	§A.4	Silent	No claim issued; invulnerability secured by having nothing to be vulnerable with
Moral dumbfounding (Haidt): verdicts without reasons	§A.9	Sorted	Sorted by the informed-registration anchor (§6.5): a registered party where there is one, an orthotic residue where there is none
Externalism about reasons: the contested premise that favouring is grip	§A.5; §A.6	Costed	The content of "must", and the reference, reliability and exit costs of §§3.5–3.6 (Register, Contested premises)

The verdicts are the chapter's own vocabulary: a position is placed when the partition of §3.5 assigns it a cell and the cell's costs fall due; costed when it survives at a stated cost; absorbed when it holds the account under another name; converged when it reaches the account's

endpoint by another road; dissolved when the commitment it convicts the discourse of was never available; contained when it receives what it wanted minus the clause that says nothing; a tolled exit when it survives the sorting unrefuted at a cost itemised where it is counted; without purchase when what it would erode is the account's ontology and no semantic construction; sorted when its cases divide by an instrument the account already holds; and silent when it issues no claim the account could meet.

A.1 The naturalist absorption

The trilemma's target has been the non-naturalist; a different realist survives it. For the naturalist (Railton, 1986), badness is no inert addition but an ordinary natural fact — badness just is suffering, as water is H_2O — with emotion as its fallible detector rather than its constituent. Her property causes tears, revolts, and therapy, so the inertness on which the trilemma and the vacancy result both turn gives no purchase on her. She must be met separately, and the meeting is a fork.

Ask what the identified fact is. If badness is suffering, look at what suffering is: not neutral tissue damage plus a value, but a felt state with the against-ness built in — negative valence bound to content, the directional difference of Chapter 13 under its ordinary name. The identification "badness = suffering" is therefore the identification "badness = an oriented response", which is the present account wearing a different jacket.

The naturalist has not located value outside the responses. She has located it in the affected party's response and declined the vocabulary, so on that reading the disagreement is verbal, since every fact, every verdict, and every mechanism is shared. This is the first horn; it is absorption rather than defeat: the anchoring of §6.4, the patient's registration as the fact the observer's response answers to, just is her view, correctly described.

If instead she insists that badness is something suffering has — a further property riding on the felt awfulness, present even where the

feeling is subtracted — the second horn opens. Subtract the feeling and audit what the residue does. The tears, the flinching, the revolutions are caused by the suffering-as-felt. The postulated extra causes nothing beyond what its bearer already causes, explains nothing its bearer does not explain, and marks no difference between worlds that share their feelings. It is the inert favouring of §3.5 relocated one level down, so the trilemma and the vacancy result apply to it unmodified.

The efficacious variant. A neighbouring figure declines the identity claim and the inertness concession together: value as a property that disposes desire (Oddie, 2005), or as normative and causally efficacious through the natural states that realise it (Wedgwood, 2007). Against him the vacancy result claims nothing, since his worlds differ by the causal route — the property causes responses its absence would not — so the meeting point is this fork, with one audit added.

Ask what does the causing. The felt state and its physical base cause the tears, the flinching, and the therapy on every party's accounting, so the question is what causal work remains for the value above its base. Locate the property in the oriented response, and the first horn has him: absorption, the account under another vocabulary. Posit productive work the base does not already do, and the position secures causal autonomy for a normative property — a cost the exclusion argument collects on (Kim, 1998), itself a contested instrument, deployed here with its status in view. The figure is placed and costed. His escape from vacancy is real, at a cost in causal metaphysics, and the reliability question of §3.5 — how the caused responses come to track the property over its base — follows him through the exit.

The explanatory-kind variant. A third figure declines both the identity claim and the residue. For the Cornell realists (Boyd, 1988; Brink, 1989; Sturgeon, 1985), a moral property is a natural kind individuated by its explanatory role — a homeostatic cluster of conditions that co-occur and explain, as "health" does, without being identical to any one of them or to the responses they provoke. The kind is neither suffering nor suffering plus a value; it is not obviously inert, since it is posited to explain. The fork meets it as it met the others. Either the cluster includes registrations among its members — the

suffering, the recoil, the reform — in which case the kind is constituted in part by responses and the anchor of §6.4 has it, absorbed; or the cluster excludes them, in which case what the kind is invoked to explain is explained by the registrations it excludes, leaving the kind idle over its base — the exclusion argument run above against the efficacious variant with a cluster in place of a property. What survives the fork is the naturalist's own instrument — explanation — applied to the responses that were doing the explaining all along.

The life-form variant. A limiting case exposes the fork cleanly. Let the naturalist widen her fact beyond feeling — goodness as flourishing, say, since flourishing does not obviously require sentience. Is the thriving of a weed good, in a universe containing no valence anywhere? If yes, "good" has been detached from every possible mattering and names only growth, a predicate even she does not want, since it makes the flourishing of a plague equally good. If no — if flourishing counts only where some creature can register its own flourishing — then feeling was the gate all along, and the first horn has her.

A third answer has the neo-Aristotelian tradition behind it and must be enumerated rather than absorbed silently. The weed's thriving is good relative to its life-form (Foot, 2001), indexed to what the oak or the wolf characteristically does: mind-independent yet not free-floating, so the plague generalisation is blocked by the index rather than by a gate.

The reply runs the index through §2.2. Ask what life-form-relative goodness consists in. If deviation from the form is bad for the organism in a sense that anywhere counts as some creature's registration — the wolf's hunger, the plant's nothing — then feeling was the gate after all, indexed rather than removed, and the first horn has her with one extra step. If it never counts so, if the standard is pure conformity of specimen to type, gradeable in a universe of zero registration, then it is the function-talk §2.2 dissected: grading against a type is orthotic assessment (§4.3), and "a good specimen" is good in the way a correctly spelled word is — description of conformity, constituting no value.

On that reading the thriving plague is a flawless specimen, and the tradition's own recoil from calling that goodness has always been the tacit appeal to what flourishing involves for the creature living it: the first horn, under its oldest name. Every naturalist realism, pressed to specify its fact, either is the response-constitutive account under another label or posits a residue the trilemma has already costed, with the life-form variant dividing between the first option and orthotic description under pressure of the same question.

The rival positions in the field are accordingly two, not three: the non-naturalist, placed at the cost §3.5 itemises, and the naturalist, absorbed — the life-form variant included, and the efficacious variant dividing between the absorption and the costed step above. The colour analogy, so often asked to carry the account, can now be relieved of argumentative duty. It remains the clearest available picture of a property that is response-dependent and real at once, while the warrant is the trilemma and the fork, both internal to value, neither owing anything to any verdict about colour.

A.2 The parity result

The results so far concern what a response-independent favouring would have to be. A different objection concerns what the present account fails to supply; it should be met before the field is closed, because it is the consideration a reader minded to retain realism will finally lean on. The account explains what value is; it delivers no requirement, no ought that reaches the agent whom nothing moves.

Confront the coherent outlier (§6.4): a being who registers the child's drowning with perfect accuracy and is unmoved. The account can say that the child's suffering is bad, anchored in her own registration, and that the outlier has misregistered nothing. It cannot say he is required to act in a sense that binds him independently of everything in him. Realism, the objection concludes, can — and a metaethics that cannot has explained morality's anatomy while amputating its authority.

The audit of the claimed advantage. The reply is an audit of the claimed advantage, and the audit returns nothing. Ask what the realist's requirement does when it reaches the unmoved agent. It does not restrain his hand, since no wrongdoer in the history of conduct has been halted by an ought-fact that engaged no response in him; restraint is the work of other agents' responses, of persuasion, and of the enforcement that codified responses license (Chapters 21–22). It does not enter his deliberation, since a consideration that engages no evaluative response is motivationally inert (§6.1). It does not even secure his assent, since assent is a response and he was stipulated to lack the relevant ones.

Nor is this the defect of one position. It is a structural fact about every position whatever: nothing compels assent anywhere. The book's own foundation is the limiting case, since the denial of difference instantiates difference (Chapter 1) and the denier can still say no — exhibited rather than converted. If self-vindication does not turn the resisting mind, no metaethical posit was going to.

Call this the parity result: the capacity to bind the unmoved belongs to no theory, fact, or foundation, so its absence here is the absence of something from everything, and discriminates between no positions. The logic is Chapter 20's reply to the universality objection (§5.2), run in reverse — a defect every position exhibits identically has become a condition, so an objection that charges it to one party has made an accounting error. The must the objector demands already exists in the account, in the political layer: where requirements do take hold of the unwilling, they do so as law, shared preferences codified through persuasion and backed by enforcement (Chapter 22, §5), reaching the unmoved agent through his circumstances — the only route any must has ever taken.

The objection has a second, factual form. A sophisticated realist concedes the audit — his requirement moves no one it does not move — and retreats to the fact alone: whatever anyone feels, it remains true that the outlier ought to act. The retreat surrenders the property the parity result audits and retains the clause the vacancy result empties: a requirement stripped of every engagement with any response is the inert favouring of §3.5 under a practical name, and §3.6 falls due unchanged.

The two results close as a pincer. Read the ought as gripping, and parity answers: nothing grips, the realist's included. Read it as idle, and vacancy answers: an idle ought says nothing. Grip and idleness partition the case by negation; there is no third reading.

What survives should be stated, because it is real and the account keeps it. The felt bindingness of moral requirement is a directional difference (Chapter 13), the pull of the empathic modality that the traditions misattribute to rule, procedure, or command (Chapter 16, §4.1), and for every agent who has it, morality obligates just as experienced. What the account declines is the projection of that pull into the facts (§3.5), where it would await even the agent who lacks it. The residue is §A.7's: a being altogether without the pull is silent, and from silence no objection issues, while the objector who demands to know why the silent being is nonetheless wrong exercises, in the demand, the orientation he imagines absent.

The parity result is established. It rests on one descriptive premise — assent-compulsion is exhibited by nothing — and the premise would be idle if whatever turned an agent were classified as a response by definition. The classification is empirical instead: a response is an engagement of the evaluative modality, carrying the marks this chapter has assigned it throughout — felt pull, behavioural orientation, physiological trace — each detectable independently of the conduct it explains.

The denier's counterexample accordingly has a stateable shape: an agent whose conduct or judgement alters upon presentation of the putative requirement while every such channel is shown flat, sanction and incentive included. Nothing of that shape is on record, so the premise stands as an empirical universal with its refutation condition open. The retreat position falls to §3.6. The outlier returns at §A.5, where he divides into the agent who is moved and outweighed and the agent who knows only the rule, and where the premise the objection rests on is modelled and costed.

A.3 The commitment result

The vacancy result was built against the robust realist. It has a second edge, so far unspent, and the error-theoretic tradition stands on the side it cuts. Mackie's (1977) argument from queerness runs: ordinary moral discourse is committed to objective prescriptivity, values in the fabric of the world binding independently of every response; no such properties exist; so ordinary moral claims are systematically false.

The second premise is this chapter's own §§3.5–3.6, reached by another route, so the tradition has treated error theory and response-constitution as neighbours divided by a verdict: agreed on the world, split on whether moral talk survives it. The contemporary statement sharpens the target without moving it, since Olson (2014) locates the error in a commitment to irreducibly normative favouring relations, which is the third cell named as such, and the commitment attributed is the vacant clause under its most careful description. The division is unavailable, because once the vacancy result holds the first premise cannot be formulated.

Why the first premise cannot be formulated. The first premise attributes the vacant clause to moral discourse as a truth-condition. A truth-condition is a contribution to content, a conjunct that narrows what a claim rules out, and by the biconditional of Chapter 7 (§2) a condition is the condition it is through the contrast it marks. The response-independence clause marks none (§3.6). A conjunct that excludes nothing adds no condition, so a clause that cannot narrow a content cannot figure among its truth-conditions. The other reading of the clause (§3.6) gives the error theorist a contentful commitment and takes away the attribution: read with "value" as valence, the clause says that a pull obtains with no body pulled, and moral discourse is no more committed to that than to married bachelors. The phenomenology he cites shows a direction projected into the facts (§3.5); it does not show speakers committed to the direction's subsisting unfelt, and the evidence that they are is his to produce.

The commitment Mackie's argument needs is therefore impossible rather than false: there was never a seat for the passenger. What "cruelty

is wrong" is answerable to is its contrastful remainder — the affected party's registration, the oriented responses, the facts §3.7 catalogued — and, so answerable, it is very often true. Call this the commitment result: the error theory keeps its second premise and loses its target. There is no error of the alleged kind for moral discourse to be committing, because there is no content of the alleged kind for it to be committed to.

One reply is strong and must be met. The error theorist can relocate the commitment from sentences to speakers: ordinary moralisers intend the objective prescriptivity, the phenomenology of demand confirms it, and a discourse whose users systematically intend the impossible is infected however clean its semantics.

But intending the vacant adds nothing to what is intended, for the reason that asserting it adds nothing to what is asserted: the will needs a content for its object, and §3.6 shows there is none. What the speaker has is the phenomenology, which §3.5 has already secured — the projection a mind of directional responses undergoes whether or not the clause it gestures at says anything. The error theorist reads the seeming as a commitment and convicts the discourse; the account reads it as a projection and convicts nothing, since a discourse is answerable to its contents and a seeming is no truth-condition. The choice between readings is closed, because the first requires the seeming to have a contentful object, and §3.6 stands between it and one.

Comparison with Streumer. The result should be measured against the strongest recent argument in the neighbourhood, because the two run in the same direction on different engines. Streumer (2017) argues that the error theory cannot be believed, since believing it would require believing that there is no reason to believe anything, which no believer can manage. His conclusion concerns the believer; the commitment result concerns the theory's target.

Streumer leaves the error theory standing and unbelievable, a position its defenders must inhabit without occupying. The commitment result takes the theory's first premise away: a discourse cannot be committed to a content that does not exist, so the attribution fails before

any verdict of falsity is reached, believed or not. The two are compatible, and their order matters. If the commitment result holds, Streumer's finding is a symptom under explanation — the theory resists belief because the commitment it attributes was never available to be the object of one.

Joyce's (2001) fictionalism inherits the fate at one remove. It prescribes retaining moral talk as useful pretence, and a pretence presupposes a defect: the literal content is false only if the vacant clause belongs to it, which the commitment result denies. Strip the clause and the fictionalist recommends pretending what is already true — assertion with stage directions. The utility he registered, the account keeps; the verdict of falsity that made the utility need a licence is removed.

The commitment result is established, inheriting §3.6 and the Chapter 7 biconditional and adding no premise of its own. A challenger must reopen §3.6 or reject the biconditional, which stands on the determination dilemma (Chapter 2, §2.4), which is where a rejection has to be lodged.

One dependency should be flagged, because it is unique to this result. The other results survive a weakened vacancy reading, since a critic who downgrades §3.6 from "says nothing" to "says nothing discernible" leaves the trilemma's costs and the containment accounting intact. The commitment result does not survive the downgrade: a clause that says something indiscernible could still be a truth-condition, and Mackie's attribution would be reformulable. This result alone requires strong contentlessness — the clause marks no contrast, not an inscrutable one — and the residue test of §3.6 supplies it, since a term whose only anchor the clause itself subtracts retains nothing to say indiscernibly. The Register entry records the dependency and its discharge. The result's yield is used at §A.4: both wings of the field needed the vacant clause, one to affirm and one to mourn, and the same argument retires both.

A.4 The quarantine of the field

Assembled, the results of §§3.5–3.7, A.1–A.3 support a claim none supports alone; it should be made explicitly, graded to its tier, and exposed to failure. The claim: every stated rival to response-constitution is placed, costed, absorbed, converged, dissolved, or silent, and the partition beneath it is exhaustive by derivation.

A position in the metaethics of value is individuated by where it locates the favouring value involves, and §3.5 established, by logical form, that a registered contrast is transferable (structure), proprietary and oriented (valence), or composed of no contrast (nothing). Every possible location is one of three cells; every actual position is a stance toward one of them; the field can be read off the partition.

The census by cell. The non-naturalist locates the favouring in the third cell and is placed: clause vacant, reference unfixed, reliability circular, authority binding no one (§§3.5–3.6, A.2). The cell's most careful occupants should be named rather than left to the type. Parfit's (2011) irreducibly normative facts, Scanlon's (2014) fundamental reasons, and Shafer-Landau's (2003) non-natural moral facts are third-cell claims whatever their differences, since each posits a favouring that is neither structure nor felt pull. Each meets the vacancy result at the clause and the reference argument at the hook.

The circle defence of the family's concepts, the self-standing normative vocabulary, is met at §3.6 and stands below as one of the two tolled exits. Scanlon's location of a reason's purchase in rational self-governance is met separately, at Chapter 16 (§3.1), where the seeing that governs is shown to carry the felt must or to move nothing. Gibbard's (1990) norm-expressivism lands where Blackburn's does and by the same route.

Korsgaard's (1996b) constitutivism needs its own line, since it declines the sort as stated: value is conferred by the reflective structure of agency, which is offered as neither a response nor a response-independent fact. Run the conferring through the partition. Either the reflective endorsement engages the agent — carries a felt must, is something it is like to undergo — and the second cell has it, with the

practical identity supplying the content the valence rides. Or it is the registration that a relation obtains, which is the first cell, and Enoch's (2006) shmagency question is the demand for the pull the registration lacks. The shmagency objection is met at §A.7 and counted among the batteries at §A.8; the constitutivism it targets is met here: the objection lands because the constitutivist has no third place to stand either.

The naturalist locates it in the second cell under other names and is absorbed (§A.1), the residue and life-form variants returning to the third cell. The quasi-realist declines to locate it and builds instead — attitudes stabilising into intersubjective, truth-apt discourse, an endpoint structurally identical to the account's, reached by a longer road at the cost of one idle step (§6.2). Creeping minimalism leaves the account untouched, because its distinctness is ontological (§A.8).

The error theorist and the fictionalist affirm the third cell in order to convict the discourse of it, and are dissolved (§A.3). The eliminativist denies the second cell and meets the production challenge of Chapter 13 (§2.3): he must exhibit suffering that occurs while nothing is undergone. If he argues instead, §A.7 meets him — a being altogether without the responses is silent, and from silence no objection issues.

Two engineered refusals. Two positions are engineered to refuse the sorting and must be met on their own construction. Conceptual-role realism (Eklund, 2017) is the reference relocation of §3.5 performed with finer tools: "good" names whatever property plays the action-guiding, deliberation-closing role the concept occupies in thought, so the role does the fixing and the property need only occupy it.

Run the role through the partition it was built to bypass. A role is a pattern of doings — guiding, closing, licensing — and an occupant qualifies by doing them. A structural item can occupy the role's third-person shadow, but a vector guides no deliberation and closes nothing. An inert occupant cannot be selected by a doing-role over the endless inert properties that fit the same shadow, which is the permutation problem Eklund himself presses, returned to sender. What does the guiding and closing is the felt pull in exercise (§A.7), since deliberation is closed for someone by something that moves her. Fill the role with its

only working occupant and the position has located value in the second cell: the vocabulary kept, the occupant named, the realism gone in the naming.

Quietism — the Wittgensteinian counsel, and Dworkin's (1996) argued version — holds that metaethics itself is the mistake: moral claims are to be taken at face value, and the question of what makes them true invites no answer above the moral level.

Against the argued version the containment result (§3.7) replies without combat. The face value quietism wants — truth-aptness, truth, independence of the observer's take, corrigibility — is what the account delivers item by item, with the vacant fourth clause the only subtraction. By his own lights he cannot mourn the subtraction of what says nothing. Argued quietism is the containment result practised without its accounting. Against strict refusal to engage, no argument is owed or possible: a position that declines to say what it locates where has issued no claim, and invulnerability of that kind is secured by having nothing to be vulnerable with.

The two tolled exits. Two exits stand outside the sorting, each at a stated cost, and the census closes by counting them. The primitive-structure magnetist (§3.5) secures reference without mediating content, at the cost of a primitive normative joint and the reliability circle unclosed. The self-standing-circle intuitionist (§3.6) secures his family's standing by declining the anchoring every companion family has, at the cost of a concept with no exhibition and no composition, a faculty that would reach an inert property unaffected by it, a vocabulary whose use nothing regulates, and the same circle unclosed behind it.

Neither exit holds a located content — one runs past content, the other declines its anchoring — so neither re-enters the field as a rival. Each survives as a position held at a stated cost, and each is a worked example of what remaining outside the field requires. The residue test states what either would produce to re-enter: an anchored remainder for the vocabulary, or a candidate content for the magnet to select.

What the quarantine grades. The enumeration is complete, and the grading must resist the inflation this is the moment for. It must also

mark what kind of claim each half is. That the three cells exhaust the possible locations of a favouring is established, by logical form, on a chain with no constructed link. That every position actually advanced occupies one of them is a survey, and a survey is defeasible by a position it has not met. A challenger who finds one corrects the enumeration and leaves the partition standing, owing, at the next step, the location the partition demands of it. That none currently does is reported rather than derived, where the report is the half that a better-read critic can improve.

What the quarantine establishes is negative: no contentful rival occupies the field. That result rests on a chain with no constructed link — Chapter 1's mechanism, the exhibitive closure of Chapter 7, the partition of §3.5 — together with the one commitment §3.6 states, so a challenger must reject a defended premise by name or refuse that commitment at its stated cost.

One such premise should be named here, because it is the one the externalist about reasons will reject: that a favouring which reaches no register, gripping nothing and marking no contrast anywhere, is no favouring (§3.5, §A.2). The externalist holds that reasons apply without bearing on any agent, and rejects it. The Register carries the rejection as a contested premise with its cost attached — the reliability circle of §3.5 unclosed, the reference route of §3.6 unanchored. §A.5 then states the premise as a model of "must" the externalist cannot refuse without paying for the word.

Everything positive remains constructed: that the second cell's occupant is the evaluative modality of Chapters 13–14, that feltness is availability (Chapter 13), that the moral species of the response is empathic (Chapter 16). Each has its own exposure, recorded in its own Register entry, and none is strengthened by the quarantine.

A reader may still refuse the account. She may take one of the two tolled exits, at the costs attached; she can no longer refuse it for a rival that has not met its cost. The refusal that remains below the exits is refusal of the chain, whose floor is the determination dilemma of

Chapter 2 (§2.4), where the cost of refusal is contents that cannot make any thought the thought it is.

What she holds afterwards is the residue §3.5 predicted: the report that the vacant clause seems meaningful to her, a seeming the account explains and cannot argue away, since arguments operate on the pulled. The residue is logged in the Register and the General Conclusion at its width — one account, two tolled exits, one report with no content located for it to be right about, and one named premise, that favouring is grip, whose rejection is the externalist's door. §A.5 stands on the far side of it.

A.5 The field and the pull

The premise just named — that a favouring which reaches no register is no favouring — is the one the externalist about reasons refuses. So far the refusal has been treated as a door: a named premise, a stated cost, and a reader free to walk through. That is correct as bookkeeping and incomplete as argument, since two parties who each refuse the other's first premise have not been adjudicated by either's naming it.

This section supplies the adjudication that is available. It does not compel the externalist; nothing in this dispute compels anyone. It shows three things: that the premise he refuses is one he uses everywhere else; that refusing it here costs him the content of his own vocabulary; and that what he buys with the refusal does no work. The shape is the burden asymmetry of Chapter 11, and the model that carries it is one both parties already accept.

The model. Gravity must pull. Take the "must" apart. The earth's field is present whether or not a body is in it; the pull is present only when a body is. "Must pull" means that wherever a body is placed, the pull follows. It never means that the field pulls when there is nothing to pull. Field without body is a shape: a structural fact, there to be discovered, and inert. Pull is field together with body.

"Field" here is the physicist's word, the region in which a body is pulled; it is distinct from the evaluative field of Chapter 15, the set of

pulls on one agent; where both are in play below, the second is called the evaluative field.

Now put a reason through the same model. The fact — the child is drowning, the dog is suffering — is the field, and the account grants that it is there whether or not anyone registers it; that is the anchor's observer-independence (§6.4). A register is a body: something in which a tendency can show up. The reason is the pull. One restriction keeps the model from proving too much: the bodies that reasons pull are registers of one kind, evaluative modalities, felt direction in the sense Chapter 13 fixes, and not every system with a directed response, or the thermostat would have reasons. With that restriction the model says what the premise said, in a vocabulary the externalist cannot disown — a reason for an agent is a pull on that agent, and a pull needs a body.

The classification. Ask what a reason would be that existed before any register. A circle drawn in sand is the letter "o" only inside an alphabet that gives the shape a role: a place among the other letters it contrasts with, a way of being read, a way of being written wrongly. Before alphabets it is a circle — present, round, and no letter, because there is nothing for it to be a letter in. A scribble that looks like a letter looks like one only to a reader who already has the alphabet.

So with the reason. Either it is a shape: a structural fact, there before any reader, where numbers also live, and primeness does not tell seven what to do. Or it is a role: a standing something has within the practice of being taken up, acted on, and got wrong, which needs the practice. The externalist's comparison of normative facts with mathematical ones lands his reason on the shape side, where nothing favours. What he wants is a circle that is already the letter before any alphabet exists, something that pre-exists every register and applies to someone. That is the third cell of §3.5, drawn so that its occupant can be seen.

The jobs. What the externalist hires response-independence to do can be listed, and the list checked against the anchor.

The job	How the account does it	Where
A person can be wrong about value	Her registration misreads, or reads, the affected party's	§6.4
A whole society can be wrong	The fact sits in the victims' registers, and the informed-registration anchor reaches victims whose own registration has been bent	§6.5
Moral views improve rather than merely change	Correction of registration	Chapter 16, §6.7
What the unmoved do is wrong	Settled in the victim's register	§6.4
The unmoved are criticisable	By the registers of those who respond to them as sources of harm, and by the rule	§6.4; Chapter 16

The one sentence left to the externalist is "he has a reason of his own not to". On the model it has five readings — the field, the rule, the sanction, a masked tendency in his own register that correction would release, and our criticism of him — and each is a job already done. A sixth reading, doing work none of the five does, has been sought and not found. The Register records the search as open, and the challenge as the externalist's to meet.

The outlier of §A.2 divides into two, and the division matters: the common case is the account's, and only the rare case is a commitment. In the common case the agent feels the pull, knows the act is wrong, and does it anyway, because anger, appetite, or fear pulls harder. The reason is a pull in his evaluative field, and it loses the resultant. That is what

Chapters 15 and 18 say about every choice — forces with direction and magnitude, a composition rule, the act following the strongest total — and it is what weakness of will has always looked like from inside. The externalist's "he knew and was unmoved" is false of this man. He was moved, and outweighed.

The two cases of the unmoved. The rare case is the agent with no empathic modality. He knows the act is wrong only as one knows a rule: people call it wrong, it is forbidden, it is punished. That is orthotic knowledge — correct classification under a standard with no registration beneath it — and the clinical literature has a phrase for it: the psychopath knows the words but not the music (Cleckley, 1941/1976; Johns & Quay, 1962). He holds the word "wrong" with its inferential connections and none of its content, the self-standing circle of §A.4 in the flesh.

For him the account says five things. His act is wrong, settled by the victim. The consideration against it is real, and inert for him. If he judges the act fine, he is mistaken about the standard's application. He is criticisable, and subject to the rule and its sanction. And he is not irrational; he is modality-reduced (Chapter 16, §5.2). The commitment is that last sentence; it is smaller than the objection made it sound: the reason is there, and inert for him, and the intuition that it still has hold of him is our pull, projected into a register that has none. A third figure, whose modality is intact yet at rest while he judges, is the common case; §A.6 places him.

The costs, unified. Three costs were charged to the third cell at §§3.5–3.6: reference unfixed, reliability circular, vocabulary unanchored. The model shows them to be one. A force is individuated by what it does to bodies: two forces that do nothing to anything are not two forces. It is detected by what it does to bodies: a force that moves nothing cannot be found out about. It is understood by what it does to bodies: "must" bottoms out in pulls, and a "must" that bottoms out nowhere has the word and not the content. A reason that pulls no body fails all three at once. The externalist's payments fail the same way. The pre-arranged harmony by which evolution is said to have tuned our responses to the facts (Enoch, 2011) needs the facts to have done

something for selection to track, which makes them a field with bodies in it. The companionship of mathematics places his relation on the shape side, where the successor function is individuated by its structural role and tells no one what to do.

Epistemic reasons. The sharpest form of the reply is that the account uses what it denies. There is a reason to believe what the evidence supports; it applies to the dogmatist who is not moved by evidence; and this book invokes such reasons each time it says a challenger must reject a premise by name (Cuneo, 2007).

Run the model. The evidence is the field; the truth-seeking register is the body; the reason to believe is the pull. The dogmatist's belief is incorrect, a structural fact about its fit with the evidence, and the dogmatist is unbound, since nothing in him takes the evidence up. Correctness is structural and bindingness felt. The book's own arguments stand under the same description: correct, if the premises hold, and binding on the reader they move. That is the parity result applied to the book itself (§A.2), and it exempts nothing.

Three last resorts. Three last resorts remain to the externalist, and each lands in a cell. The first is the world with no register in it: the last person destroys the forest, no one will ever know, and the externalist says "still wrong". The case is never judged from nowhere. Whoever says "still wrong" is a body in the field; the saying is the pull. Subtract every body, the judge included, and what remains is the field, which is there and favours nothing. The account keeps "idle" for a natural fact no one takes up and "vacant" for the normative item said to sit beside it.

The second is deliberation (Enoch, 2011, ch. 3). When an agent asks what she should do, she is not asking what she is pulled toward, which she already knows, so the question is said to presuppose an answer beyond her pulls. It presupposes an answer beyond her present pulls: the field she would have if it were enriched and informed (Chapter 15, §4.4), and the gap she feels is the gap between that field and this one.

The third is the question behind the question: why should she care about her own pulls at all? Caring about caring is a pull; the regress ends where the pulls end. There is no standpoint outside the field from

which the question is put, and the dissent that seems to come from outside is one more vector in it.

What the section delivers. The quarantine of §A.4 was stated with a premise the externalist refuses, and a refused premise is a standoff. The field and the pull convert the standoff into an asymmetry of the kind Chapter 11 charged the kalām. The account does every job on the list with no posit. The externalist's posit does no further job, has no body, and carries one unpaid account: the content of "must". And the one commitment the account makes — that the modality-reduced agent who knows only the rule is not irrational — is stated at its size and costs less than it sounded.

The externalist may still refuse the model. What he cannot do is refuse it for free, because the word he keeps is the word the model gave content to, so the burden of saying what else it could mean is his. That is the grade the quarantine carries: every rival placed and costed, with the burden on the refuser. One demand of his remains, and it belongs to the next chapter — that there be a fact about the hard trade-off, where two registrations conflict. Chapter 15 (§4.4) takes it up, taking the model with it.

A.6 Representation and registration

The two cases of the unmoved left the common case out. Between the agent who feels the pull and the agent who has never felt it stands a third, who has felt it, holds the verdict, yet at this moment feels no pull at all: the reasoner who says of torture, without a flicker, that it is wrong, reporting no emotion in the saying. He is the common case — every competent moral reasoner in a cool hour, the realist philosopher at his desk among them — so his report is the datum the realist's seeming rests on: apprehension with no affect in it. §3.5 explained why value presents as response-independent whenever it presents; it did not explain how a verdict can be held while no response occurs. This section supplies the mechanism, concedes the report whole, then shows that the concession costs the account no result.

The three cases. An agent stands to a moral verdict in one of three ways, sorted by the history of his register together with its present activity. In the first the verdict was never anchored: he holds the word with its inferential connections, having registered no pull beneath it at any time — the modality-reduced agent of §A.5, who knows the words but not the music. In the second the verdict was anchored while the register is now silent: a pull was registered, the direction it carried passed into a representation, the representation is deployed while the register rests. In the third the register is engaged as the verdict is made: the pull is undergone, the verdict its representation. The first case the model placed as the rare one, the third as the ordinary run of moral life. The second is where the realist stands when he reports seeing without feeling, so it is the case the account must place without denying the report or granting the seeing a third source.

The mechanism. The anchoring requirement fixes the direction of content between a label and a registration: labels transmit content, registration supplies it, the containment does not reverse (§3.6). What the requirement leaves open is the tense. Supply need not be simultaneous with transmission, since in the foundational vocabularies it never is: the child who learned “two” from two apples counts without apples for the rest of her life, the numeral carrying what the apples supplied, while no one infers from her cold arithmetic a second exit into a realm of number apprehended without counting. The evaluative case takes the same tense. A pull is registered — the suffering registers as bad, the toward or the away is undergone — whereupon the direction it carried enters a representation: “counts against”, “a reason not to”, “wrong”. The direction is a grasped diff; the label attaches to it as labels attach (Chapter 7, §3). Once attached, the label runs on its own connections. It enters inferences, is weighed, applied to new cases, taught, argued over, with the register at rest, since cold deployment is what a learned vocabulary is for. The alphabet of §A.5 supplies the picture. A reader who learned the letters by sounding them reads silently for the rest of her life, the letters keeping their roles unsounded because the roles were fixed when the sounding was: no silent reader without the sounded alphabet behind her, no sounding required once it is there. The tense is not peculiar to value. Affect of every kind is recalled as

representation, since the brain re-presents the body's response without re-running it — Damasio's (1994) as-if loop — so pain is remembered as the image of pain with the pain mostly gone, and terror as the picture of the moment; the cold verdict is the ordinary tense of any remembered feeling.

The seeming. This places the realist's report without disputing it. When he says that he feels no pull while seeing that torture is wrong, both halves are true. He feels no pull at the moment, since the register is silent; he holds the verdict, since the representation is intact. What the report does not contain is what he takes it to contain: a seeing whose direction came from somewhere other than a register. The direction in "wrong" — the against-ness §3.5 showed a favouring must carry — was supplied at the anchoring, sits in the representation as a grasped diff, so it is what the cold verdict manipulates. §3.5 diagnosed why the direction presents as the world's: a mind of directional responses cannot represent a value undirected. This section adds why it presents so at the moment the register is at rest — because the representation carries the direction whether or not the register is running, so silence in the register reads, from inside, as apprehension with no response in it. The realist reads the autonomy of the representation as the autonomy of what it represents. The two are distinct on the model as everywhere: holding the record of a pull is not being pulled, as holding the field equation is not falling. One feature of the report the realist counts as evidence remains to be placed: the force of the seeming, the assertive character with which an intellectual seeming presents its content and mere supposition lacks (Huemer, 2005). On the model the force is the residue. The as-if loop seldom runs cold, so the representation arrives with a trace of the pull it records; the trace is what makes the verdict feel like perception where the same sentence merely supposed does not. The first case is where the trace should be absent, which the stake below can carry.

What is not claimed. The placement is not a genealogy. That a concept was acquired through registration settles no question about what it can be true of; the realist who relocates his claim from characterisation to reference was met at §3.5 on other grounds, origin not among them. What the mechanism removes is narrower, while

sufficient for its purpose: the evidential weight of the cold report. The realist offered “I see it without feeling it” as a datum his opponent must explain away or else concede a third source. It is explained, on the account’s own resources, with no third source, so the datum discriminates his view from this one no better than the seeming of §3.5 did. The refuting is done elsewhere, by the vacancy result on the clause with the trilemma on the favouring; this section leaves both where they were, taking from the realist one thing only — the phenomenology he had been counting as evidence.

The burden. The sixth-job search of §A.5 can now be put in the terms the realist prefers. Two things are present in any cold moral verdict: a representation, with the inferential connections of the vocabulary, plus the registration that anchored it, present as history where it is absent as occurrence. The realist needs a third — something in the apprehension that is neither the representation nor the anchoring registration, since the account already owns both. The Register keeps the search open with the sixth job, the burden on the same party: to exhibit, in a cold verdict, a component the two named components do not exhaust. The channel question puts the same demand in the realist’s own terms. The apprehension is in symbols, the sentence “torture is wrong” in inner speech, with an image or a memory of what is judged beneath the symbols. Symbols carry content; an image carries no direction, since the picture of the child struck contains no “to be stopped” among its shapes. So the against-ness is supplied by the image, which cannot supply it, by the pull, or by a third channel that has yet to be named — and the clinic does not find one, since removing the affective contribution changes the verdicts while the symbols and the images stay intact (Koenigs et al., 2007). “Intellectual” names no channel; it names the symbols.

The clinical face. The three cases come apart in the clinic, where the account predicts how. Patients with acquired damage to the ventromedial prefrontal cortex — “acquired sociopathy” (Damasio, 1994) — keep their moral vocabulary, deliver normal verdicts on most moral questions, yet the affect that once accompanied those verdicts is blunted, so the verdicts no longer move them to act. Roskies (2003)

reads the population as a counterexample to internalism: judgement intact, motivation gone, so judgement is separable from motivation. On the account it is the second case under damage — representation anchored before the injury, register impaired since — so the account expects the retained verdicts to fail where a fresh registration would ordinarily have decided the case. That is what the data show: patients of the same kind give abnormally 'utilitarian' verdicts on the high-conflict personal dilemmas, where the aversive pull normally settles the matter, with normal verdicts elsewhere (Koenigs et al., 2007). The developmental case differs as the first case should: where anchoring never occurred the moral/conventional distinction itself is impaired (Blair, 1995), the vocabulary learned with no diff beneath it, the words without the music. Kennett and Fine (2008) defend internalism against Roskies by questioning whether the patients' verdicts are moral judgements in more than an inverted-commas sense; the account needs no such denial. The verdicts are competent representations; what they lack is a running register, which the account had already separated from them.

The stake. Stated plainly, the two populations should diverge on a test the chapter specifies and none has run as such — matched for vocabulary competence, the developmental case and the acquired case compared on the moral/conventional distinction, on the sensitivity of their verdicts to the affected party, and on the force with which the verdicts are held. The account predicts the acquired case ahead on all three. Convergence, the developmental case reaching the acquired case's competence with no history of registration, would break the placement; the Register carries it among the empirical stakes.

What the section delivers. A placement rather than a result. The cold verdict is the second case, its direction supplied at anchoring and carried by the representation; the realist's report is conceded whole, explained with no third source; the sixth job is restated in his terms and left with him; the clinical dissociations of §6.4, which laid the floor under the sufferer, are joined by a dissociation on the judge's side, which the account predicts where tracking must treat it as coincidence.

The grade is that of a defeated objection. It establishes no thesis, adds no row, and leaves the load where §3.6 and §A.4 carry it.

Where the argument stands: value has been located in the second cell and the field quarantined (§A.4), with the premise beneath the quarantine modelled and costed against the externalist (§A.5), with the cold verdict placed as the second of three cases (§A.6). What remains are three objections the account's own construction invites (§§A.7–A.9).

A.7 The self-application objection

The objection is that the account universalises: all value is response-constituted. The value of believing what is true is then response-constituted too, and with it the force of this book's own arguments, including the arguments for this account. When the chapter concludes that a reader ought to accept the trilemma's result, that "ought" is, by the account's own lights, a felt pull rather than a binding fact, so a reader unmoved by it has, on the account's own showing, made no mistake that obtains apart from someone's response. The framework appears to saw off the branch it argues from.

In reply, the account accepts the description and denies that it is a wound. Two things must be separated, as this chapter separated them for moral verdicts: the correctness of a derivation and the force of its conclusion.

Correctness is structural. A derivation is valid or invalid in virtue of contrastive relations among contents (Chapter 7, §2), obtaining whether or not anyone is pulled by it — response-independent in the only sense the objection needs, which is all the Register's "established" gradings record. What is response-constituted is the force: the pull toward believing what follows, the mattering of getting it right. That pull is an evaluative response like any other, an alethic orientation, valence coupled to registrations of accuracy and error.

But this particular response occupies the dialectical position the diff occupied in Chapter 1: it cannot be disputed without being deployed. The objector who argues that the book's conclusions do not bind him is

arguing — advancing considerations because they are correct, expecting them to be weighed because getting it right matters — so every move in the dispute exercises the alethic pull whose authority the dispute denies.

The force of the book's arguments is therefore response-constituted and inescapable for any participant in argument. Not because the framework compels the orientation, since consistent with §3.5 it obliges no one, but because entering the practice of argument just is deploying it. A being altogether without the alethic pull is not refuted; he is silent, and from silence no objection issues. The self-application is accordingly the epistemological twin of Chapter 1's self-vindication rather than an embarrassment: there, the denial of difference instantiated difference; here, the dispute of alethic force exercises it. What the account gives up — a bindingness that would reach even the silent — is something no theory has ever had and argument has never needed.

The objection has a general form, pressed as a companions-in-guilt argument for realism (Cuneo, 2007). Epistemic normativity — that one ought to believe what the evidence supports — is categorical if anything is, so a metaethic that states the cost of response-independent moral requirements must state the cost of epistemic ones identically, and the cost there looks unpayable, since inquiry itself would be forfeit. The account meets it with the resource just displayed, and inquiry stands.

Epistemic assessment divides as moral assessment divided (§4.3). The orthotic component — a belief's conformity to its evidence, a derivation's validity — is structural, obtains whatever anyone feels, and of itself requires nothing. The binding component — the pull toward the supported conclusion, the mattering of accuracy — is the alethic orientation named above, valence coupled to registrations of accuracy and error, present in every inquirer who cares whether she is right.

The companions argument shows the two domains stand or fall together, and the account agrees: both fall the same way, into an orthotic layer that classifies and a felt layer that binds, with no third ingredient in either. The guilt is shared and the guilt is costed. The epistemic ought that would await the inquirer whom accuracy leaves cold is the parity result's subject (§A.2) under an epistemic name, while the practice of

inquiry — assertion, correction, retraction — draws on registrations and responses, never on it.

A.8 The standard batteries

The account should close by passing the tests that have organised the metaethical field. The Frege–Geach problem (Geach, 1965) does not arise, because the account is descriptivist. “Cruelty is wrong” states a response-dependent fact (§6.2), so embedding, negation, and inference are the ordinary logic of statements.

Moral twin earth (Horgan & Timmons, 1991) defeats synthetic naturalism by imagining communities whose term “good” tracks different natural properties while their disagreement remains real. Here the shared subject matter is structural: oriented responses, the moral instances anchored in affected parties’ registrations (§6.4). The dispute is therefore over the calibration of one modality with one constitutive role — real disagreement without a shared natural referent, the profile §A.1 predicts.

Creeping minimalism (Dreier, 2004) erodes positions whose realist idioms are earned by construction; nothing here is earned by construction. The truth of “her suffering is bad” is the obtaining of a response-dependent fact, located in her. The position is distinguished from expressivism by its ontology and from robust realism by the vacancy result (§3.6). Street’s (2006) Darwinian dilemma finds no target, since nothing is tracked (§6.2). Enoch’s (2006) shmagency objection is met at §A.7: the only escape from constitutivism about content is silence.

Each test named a structural defect — embedding, reference, distinctness, tracking, escape — and the account passes each because the defect’s site does not exist in its architecture. A framework should be permitted one sentence of comparison after such a run: no existing programme both grounds its semantics in a foundation established by forced choice and passes the battery whole. The conjunction, more than any single result, is the account’s claim on the field.

A.9 The dumbfounding objection

The objection is that moral dumbfounding (Haidt, 2001; Haidt, Koller, and Dias, 1993) appears to show that moral judgement is driven by responses the account itself must call unfitting. Subjects confront scenarios engineered to be harmless — the family that eats its dog after the dog is killed by a car, consensual and consequence-free taboo violations — and maintain the verdict of wrongness while conceding, under probing, that they can locate no harm.

If moral evaluation were empathic response answerable to an affected party's registration, the verdicts should extinguish when the registration is stipulated away. They persist. The data suggest the judgements are post-hoc rationalisations of calibrated aversions, which is Haidt's own reading, so an account that identifies morality with empathic evaluation inherits a practice run, on the evidence, by something else.

In reply, the account predicts the data, and predicts their internal structure, which the rationalisation reading flattens. Dumbfounded verdicts divide, and the division tracks the account's own instruments.

Some persist because the stipulated harmlessness is not believed, and should not be, since the scenarios stipulate away the silent harms §6.5 anchors. The dog case is the clearest: what subjects cannot articulate but will not surrender is that the family's relation to the dog — trust given, dependence accepted — is falsified by the meal, a betrayal structure with the registering party conveniently dead. The verdict is the informed-registration anchor working faster than the subject's theory of it. Other verdicts persist because they were never empathic at all: orthotic residue, conformity-responses to a violated standard, misread as moral verdicts by the conflation Chapter 16 diagnoses in the traditions.

The two components carry a sorting signature the account states in advance: sensitivity to relational facts. Vary the betrayal — the dog died wild and unowned, the promise was released, the party did consent with full contact — and the anchored component tracks the variation,

weakening as the relational fact weakens, while the orthotic residue is insensitive, firing on the category however the relations are set.

The manipulation data point the same way, since verdicts shift when concealed relational facts are exposed, while the diachronic asymmetry of §6.5 is the signature run over time: verdicts on the dead thin toward the promise-breaker as living registrations recede, which anchoring predicts and a flat aversion would not. What dumbfounding shows is that the modality outruns the subject's articulacy — which the account affirms, since the anchor is a registration rather than an argument — and that a portion of the practice is orthotic conformity mistaken for evaluation, which is not an objection to the account but its thesis about the traditions, returned as data.

References

Audi, R. (2004). *The good in the right: A theory of intuition and intrinsic value*. Princeton University Press.

Bain, D. (2013). What makes pains unpleasant? *Philosophical Studies*, 166(S1), 69–89.

Blackburn, S. (1993). *Essays in quasi-realism*. Oxford University Press.

Blair, R. J. R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, 57(1), 1–29.

Boyd, R. (1988). How to be a moral realist. In G. Sayre-McCord (Ed.), *Essays on moral realism* (pp. 181–228). Cornell University Press.

Brink, D. O. (1989). *Moral realism and the foundations of ethics*. Cambridge University Press.

Carnap, R. (1932/1959). The elimination of metaphysics through logical analysis of language. In *Logical positivism* (A. J. Ayer, Ed., pp. 60–81). Free Press.

Clarke-Doane, J. (2020). *Morality and mathematics*. Oxford University Press.

Cleckley, H. (1941/1976). *The mask of sanity* (5th ed.). Mosby.

Corns, J. (2014). The inadequacy of unitary characterizations of pain. *Philosophical Studies*, 169(3), 355–378.

Cuneo, T. (2007). *The normative web: An argument for moral realism*. Oxford University Press.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. Putnam.

D'Arms, J., & Jacobson, D. (2000). Sentiment and value. *Ethics*, 110(4), 722–748.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. Oxford University Press.

Deonna, J. A., & Teroni, F. (2012). *The emotions: A philosophical introduction*. Routledge.

Döring, S. A. (2007). Seeing what to do: Affective perception and rational motivation. *Dialectica*, 61(3), 363–394.

Dreier, J. (2004). Meta-ethics and the problem of creeping minimalism. *Philosophical Perspectives*, 18(1), 23–44.

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. MIT Press.

Dworkin, R. (1996). Objectivity and truth: You'd better believe it. *Philosophy & Public Affairs*, 25(2), 87–139.

Eklund, M. (2017). *Choosing normative concepts*. Oxford University Press.

Enoch, D. (2006). Agency, shmagency: Why normativity won't come from what is constitutive of action. *Philosophical Review*, 115(2), 169–198.

Enoch, D. (2011). *Taking morality seriously: A defense of robust realism*. Oxford University Press.

Fine, K. (1994). Essence and modality: The second Philosophical Perspectives lecture. In J. E. Tomberlin (Ed.), *Philosophical perspectives 8: Logic and language* (pp. 1–16). Ridgeview Publishing Company.

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. MIT Press.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford University Press.

Foot, P. (2001). *Natural goodness*. Clarendon Press.

Geach, P. T. (1965). Assertion. *Philosophical Review*, 74(4), 449–465.

Gibbard, A. (1990). *Wise choices, apt feelings: A theory of normative judgment*. Harvard University Press.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834.

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, 65(4), 613–628.

Hardin, C. L. (1988). *Color for philosophers: Unweaving the rainbow*. Hackett.

Horgan, T., & Timmons, M. (1991). New wave moral realism meets moral twin earth. *Journal of Philosophical Research*, 16, 447–465.

Huemer, M. (2005). *Ethical intuitionism*. Palgrave Macmillan.

Hume, D. (1751/1998). *An enquiry concerning the principles of morals* (T. L. Beauchamp, Ed.). Oxford University Press.

James, W. (1884). What is an emotion? *Mind*, 9(34), 188–205.

Johns, J. H., & Quay, H. C. (1962). The effect of social reward on verbal conditioning in psychopathic and neurotic military offenders. *Journal of Consulting Psychology*, 26(3), 217–220.

Joyce, R. (2001). *The myth of morality*. Cambridge University Press.
Kennett, J., & Fine, C. (2008). Internalism and the evidence from psychopaths and 'acquired sociopaths'. In W. Sinnott-Armstrong (Ed.), *Moral psychology: The neuroscience of morality* (Vol. 3, pp. 173–190). MIT Press.
Kim, J. (1998). *Mind in a physical world: An essay on the mind-body problem and mental causation*. MIT Press.
Klein, C. (2015). *What the body commands: The imperative theory of pain*. MIT Press.
Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, 446(7138), 908–911.
Korsgaard, C. M. (1996b). *The sources of normativity*. Cambridge University Press.
Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy*, 61(4), 343–377.
Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. Penguin.
McDowell, J. (1985). Values and secondary qualities. In T. Honderich (Ed.), *Morality and objectivity: A tribute to J. L. Mackie* (pp. 110–129). Routledge.
Millikan, R. G. (1984). *Language, thought, and other biological categories: New foundations for realism*. MIT Press.
Moore, G. E. (1903). *Principia ethica*. Cambridge University Press.
Nussbaum, M. C. (2001). *Upheavals of thought: The intelligence of emotions*. Cambridge University Press.
Oddie, G. (2005). *Value, reality, and desire*. Clarendon Press.
Olson, J. (2014). *Moral error theory: History, critique, defence*. Oxford University Press.
Parfit, D. (2011). *On what matters* (Vols. 1–2). Oxford University Press.
Prinz, J. J. (2007). *The emotional construction of morals*. Oxford University Press.
Putnam, H. (1975). The meaning of 'meaning'. In *Mind, language and reality: Philosophical papers* (Vol. 2, pp. 215–271). Cambridge University Press.
Rabinowicz, W., & Rønnow-Rasmussen, T. (2004). The strike of the demon: On fitting pro-attitudes and value. *Ethics*, 114(3), 391–423.
Railton, P. (1986). Moral realism. *Philosophical Review, 95*(2), 163–207.
Roberson, D., Davies, I. R. L., & Davidoff, J. (2000). Colour categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, 129(3), 369–398.
Rosenthal, D. M. (2005). *Consciousness and mind*. Clarendon Press.
Roskies, A. (2003). Are ethical judgments intrinsically motivational? Lessons from 'acquired sociopathy'. *Philosophical Psychology*, 16(1), 51–66.
Scanlon, T. M. (1998). *What we owe to each other*. Harvard University Press.

Scanlon, T. M. (2014). *Being realistic about reasons*. Oxford University Press.
Schroeder, M. (2007). *Slaves of the passions*. Oxford University Press.
Schwartz, S. H. (2012). An overview of the Schwartz theory of basic values. *Online Readings in Psychology and Culture*, 2(1), 1–21.
Schwitzgebel, E. (2008). The unreliability of naive introspection. *Philosophical Review*, 117(2), 245–273.
Shafer-Landau, R. (2003). *Moral realism: A defence*. Oxford University Press.
Sider, T. (2011). *Writing the book of the world*. Oxford University Press.
Smith, M. (1994). *The moral problem*. Blackwell.
Solomon, R. C. (1976). *The passions: Emotions and the meaning of life*. Doubleday.
Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies*, 127(1), 109–166.
Streumer, B. (2017). *Unbelievable errors: An error theory about all normative judgements*. Oxford University Press.
Sturgeon, N. L. (1985). Moral explanations. In D. Copp & D. Zimmerman (Eds.), *Morality, reason and truth* (pp. 49–78). Rowman & Allanheld.
Tappolet, C. (2016). *Emotions, values, and agency*. Oxford University Press.
Taurek, J. M. (1977). Should the numbers count? *Philosophy & Public Affairs*, 6(4), 293–316.
Taylor, G. J., Bagby, R. M., & Parker, J. D. A. (1997). *Disorders of affect regulation: Alexithymia in medical and psychiatric illness*. Cambridge University Press.
Wedgwood, R. (2007). *The nature of normativity*. Clarendon Press.
Wiggins, D. (1987). A sensible subjectivism? In *Needs, values, truth: Essays in the philosophy of value* (pp. 185–214). Blackwell.
Williams, B. (1981). Internal and external reasons. In *Moral luck: Philosophical papers 1973–1980* (pp. 101–113). Cambridge University Press.

CHAPTER 15

Evaluation and Choice

Presupposes: the three-cell partition of mental features (Ch. 13, §2.3); the response-constitutive account of value (Ch. 14).

Establishes: the primary/added value distinction, individuated by the severing test; the structure of evaluation and choice; the constitution of the evaluative hierarchy; weighing as field-enrichment, and across evaluators as convergence under mutual registration; the account's convergence with rival theories and its answers to the standing challenges. All results are constructed; note 1 records the grades and the three empirical stakes.[1]

In brief. Why does a shirt your father once wore feel different from an identical shirt bought yesterday? The difference is in you — and this chapter supplies its mechanics. Total value is a composite: primary value (the direct pull of the thing itself) plus added value flowing backward from its source or forward from its effects. The instrument that keeps the distinction disciplined is the severing test: cut the association while holding the object fixed, and whatever dies was never the object's own — peel the label from the wine and the provenance-glow goes with it while the flavour's pull remains. The same mechanics answer the hardest question a response-based theory faces: when accurate responses conflict, there is no cosmic exchange rate waiting to be found. Deliberation is not measurement that fails for want of a unit; it is the widening of the field until it ceases to be tied — and between people, each party's taking in of the other's registration, with what such fields converge on the only fact a trade-off has.

---

1 Introduction

Why does a shirt your father once wore feel different from an identical shirt bought yesterday? Both are cloth; both serve the same function. The difference is in you — in the web of emotional associations that transforms a piece of fabric into something irreplaceable. The previous chapter established that emotion constitutes evaluative difference.

It did not explain how evaluation works in practice: how multiple responses with different intensities are resolved, how the result becomes a choice, and why people rank the same things so differently.

This chapter supplies the mechanics. Total value is a composite with a specific structure: primary value, the modality's direct response to the object, plus added value in two species — derived value flowing backward from the object's source, consequential value flowing forward from its effects.

The structure explains, within one mechanism, phenomena the existing taxonomies treat piecemeal: why identical objects can differ in value, why value changes without the object changing, why adaptive preferences have the profile they do, and, in the section this chapter adds to the sequence (§4.4), how deliberation escapes deadlock when accurate responses conflict. Beside what it supplies, the chapter returns no fact about the right trade-off between parties, only about what mutually informed fields settle on (§4.4), a limit every later chapter carries. The partition is not merely taxonomic. It is the mechanism by which the evaluative hierarchy is constituted and transformed, and every subsequent chapter presupposes it.

2 The Structure of Evaluation

2.1 From stimulus to evaluative response

Evaluation has a determinate structure, however rapidly it unfolds. A stimulus is detected by the sensory modalities. The detected distinctions are processed cognitively: the object identified, classified, and associated with stored representations, each association carrying its own possible emotional charge. The evaluative modality then responds to the resulting complex with attraction or repulsion. That response is the evaluative difference.

The process scales with the field. A simple evaluation of a shirt might involve liking its colour, the modality responding to a single sensory feature. A richer one adds the memory of your father wearing such a shirt: warmth, nostalgia, the pang of loss. Richer still, the

knowledge that it was produced by exploitative labour triggers a negative empathic response to the workers' suffering. The object has not changed; the evaluative field in which it is embedded has, so the overall evaluation is the resultant of all the oriented responses the field elicits.

Two features matter for everything that follows. The responses are constitutive rather than representational: the modality does not discover that the shirt is valuable and report its finding, but constitutes the value through its response. And the cognitive and evaluative operations are collaborative while distinct. A system with perfect cognitive processing and no evaluative modality would identify every association and register every fact without experiencing the shirt as attractive, repulsive, meaningful, or worthless.

2.2 Primary and secondary objects

The elements of an evaluative process divide in two. The primary object is what initiates the process, the shirt itself. The secondary objects are everything associated with it: the memory of your father, the brand's labour practices, the friend who recommended the shop. Each may carry its own evaluative weight, so what we call "the value of" a thing is rarely a single response. It is a composite, the resultant of the modality's oriented responses to the primary object and its secondary web. The distinction is structural rather than normative: in many cases the secondary objects dominate entirely, as when a plain envelope carrying a letter from a dying friend is evaluated almost wholly through what it carries.

3 Primary Value and Added Value

3.1 Primary (essential) value

The primary or essential value of an object is the emotional response elicited directly by the primary object, without association with any other elements — what the object is worth to you on its own, stripped of its web. Primary value is the baseline from which the hierarchy is built, and often the first response to register.

But immediacy is not primacy. Kahneman's (2011) "System 1" is defined by speed, not by which route the response travels: a fast response may be primary, the direct pull of a colour, or added, the automatic warmth an association discharges before any deliberation. Primary value names a route, System 1 a tempo, and the two axes cross — a point §6.1 uses.

3.2 Added (secondary) value and the severing test

Added value is any emotional response elicited by secondary objects, dividing into two kinds by direction. Derived value flows backward along the causal chain, from source to object: the shirt given by your father, the painting reattributed to Vermeer, the object unchanged while its derived value is transformed. Consequential value flows forward, from object to effects: the shirt that protects sunburnt skin, the canvas a billionaire will pay a fortune for.

The taxonomy needs an individuation criterion, because without one the primary/added line can be redrawn to suit any diagnosis. The Vermeer case invites the challenge directly: is the collector's surge of feeling added value on the painting, or the primary value of a different object, the-painting-as-a-Vermeer? If redescription can convert any added value into primary value, the distinction does no work.

The severing test. The criterion is counterfactual. A component of the total pull is added if it would not survive the severing of the association with the object held fixed, and primary if it would. Sever the attribution — same canvas, same pigments, the Vermeer identification withdrawn — and the surge collapses, so the component was derived. Peel the label from the wine and the provenance-glow goes with it while the flavour's own pull remains. The test is applied to components under intervention, which is what blocks the redescription manoeuvre: the objector may call the object what he pleases, but the intervention holds the object's intrinsic features fixed and cuts only the association, and what the cut kills was never the object's own.

The transformation caveat of §3.3 marks the test's one refinement, which must be stated with its discipline, or the caveat becomes an alibi.

The test is time-indexed: a component is added at a time if severing at that time kills it, primary at a time if it survives.

Transformation is accordingly a historical claim rather than a synchronic one. Where an association has retrained the primary response, severing no longer reverses the pull, so the test correctly reports a conversion of added into primary value. But to say so is to say the component was severable earlier and is not severable now, the conversion mediated by a traceable associative history. A transformation invoked with no history is inadmissible, since a component that was never severable is primary throughout rather than converted. This is what keeps the caveat from excusing every unwelcome verdict: it predicts a profile — severability decaying over exposure — distinguishable from primacy, which is severability's constant absence.

Stated as evidence, the admissibility conditions are three: a record of the exposure through which the conversion is claimed to have run; a dose profile, severability declining as the exposure accumulates; and the asymmetry the mechanism requires, conversion running forward with exposure and never reversing spontaneously in its absence. A transformation claim that can produce none of the three is rejected at the door.

So disciplined, the caveat is the account's theory of evaluative learning: the mechanism by which taste is formed and, run deliberately from outside, the mechanism marketing exploits (§6.1, §7.2). The time-indexing makes "primary" a stage-relative status, which the account accepts without strain. A pristine, history-free primary value would be the pre-social standpoint §6.6 disowns, and the framework's primary value was always the response's current direct route rather than its innocence.

3.3 The structural consequence

Total value is the composite of primary and added value, with one caveat: the relation is not always additive. Added value can transform the primary response rather than combining with it — learning that a

painting was made by a child prodigy may change the aesthetic response itself — so the formula is a simplification rather than an independence claim.

The consequence is striking. An object with negligible primary value can carry enormous total value: the painful procedure embraced for the life it saves, the worthless-looking object worth millions as a lost masterwork. The converse holds too, since the beautiful shirt made by child labour and the delicious food known to be poisonous are evaluated negatively overall. Added value overrides primary value and reverses rankings. This interaction makes the evaluative hierarchy dynamic, a constantly updated composite responsive to everything that enters the field; the dynamism is what makes deep persuasion possible (Chapter 21), since a newly enriched field constitutes a different composite.

3.4 Relation to the intrinsic/extrinsic value debate

The primary/added distinction intersects with the intrinsic/extrinsic distinction without being it. Korsgaard (1983) prised apart two senses of "intrinsic value", value as an end (final value) and non-relational value; M. J. Zimmerman (2001) developed the second as a monadic property supervening on intrinsic features alone, a reading Korsgaard's own cases of final value borne by relational features tell against.

Primary value is the nearest analogue of Zimmerman's intrinsic value, added value of the relational kinds, with two differences that favour the present cut. First, primary value is response-dependent, its variation across evaluators no more an objection than colour's variation across visual systems (Chapter 14, §2): the structural insight preserved, the ontological commitment replaced. Second, the primary/added taxonomy is compositional where intrinsic/instrumental is static. Total value is composed in real time as secondary objects enter the field, which is why the same object changes in total value without changing intrinsically, and why intrinsically identical objects differ. §3.5 sharpens the comparison into a prediction.

3.5 From taxonomy to partition: the derivation

As presented so far the distinction is a taxonomy, and a taxonomy invites the question convergence cannot answer: why must value divide this way? It can be derived as an exhaustive partition from results already established. Given the response-constitutive account (Chapter 14, §3.5), every component of an object's total value is an oriented response to it, and valence is never free-standing, being the orientation of a content (Chapter 13, §2.3).

The derivation. Ask of any evaluative response what content it rides. Either the content confines itself to what the encounter with the object itself delivers — the diffs registered in seeing, tasting, wearing, being with it — or it includes represented items the encounter does not deliver: source, history, effects, price, owner. Inclusion is binary; there is no third condition to be in. The first route constitutes primary value, which is why §3.1 could define it as the response "without association with any other elements", isolation being the encounter without importation. The second constitutes added value, its species falling out by the direction of the imported relation: backward yields derived, forward consequential.

One clarification guards the edge. Encounter-delivered character includes the relational diffs the encounter itself presents — the painting's contrast with its neighbours, the shirt's fit on this body — so the line runs between the delivered and the imported, cutting across the line between the intrinsic and the relational. The partition is exhaustive by its own logical form, which is what it rests on: any evaluative content either exceeds the object's registered character or does not. The diversity/equality partition divides on distinction against its absence, whereas this one does not descend from it; what the two share is the form of a contradictory pair, which is a resemblance in shape.

The derivation converts §3.4's positioning into a prediction. Final/instrumental classifies the *mode* of valuing; intrinsic/extrinsic the value's *ground*; primary/added the *route* of the response's content — three cuts at three levels, so the three cross. The collector's Vermeer and the father's shirt are valued finally, through the added route: extrinsic

final value, which the standard taxonomy accommodated only after Korsgaard prised its distinctions apart, arrives here as a predicted cell rather than a curiosity.

The partition also yields a structural diagnosis of the fitting-attitudes literature's hardest internal problem. In the demon case (Rabinowicz and Rønnow-Rasmussen, 2004), there is excellent reason to admire the saucer of mud, yet the mud is not thereby valuable, so the analysis owes an account of why the demon's reasons are of the "wrong kind".

The route partition supplies one. The demon's reasons enter on the consequential channel, since their content represents the effects of admiring rather than the mud's encounter-delivered character, while admiration is a response constituted by the primary route. The demon's reasons therefore ride a route that cannot constitute the response they demand. They can make admiring useful; they cannot make it registration. Wrong-kind reasons are wrong-route reasons.

This is the diagnosis of Chapter 14 (§6.4) at a finer grain rather than a second one competing with it. There the demon's reason was placed by axis, bearing on the value of hosting the response rather than on the accuracy of its content. Here the same reason is placed by route, entering on the consequential channel of a response the primary route constitutes. The axis analysis says what the demon's reason fails to touch; the route analysis says by which channel it arrives. Both are needed, and neither is prior.

A challenger to the partition has two options. She must reject response-constitution itself, or exhibit an evaluative response whose content neither confines itself to the encounter nor imports anything beyond it — composite objects included, since a configuration the encounter presents is encounter-delivered content (§7.6). The two-axes architecture leaves nothing for such a content to be. With the partition derived, the mechanics of choice can be set out.

4 The Structure of Choice

4.1 Choice as acknowledgement of preference

An act of evaluation leads to a choice: the acknowledgement of a preference between options, requiring at least two evaluative responses of differing intensities. Between two shirts, we select the one eliciting the stronger positive response, or, in an all-negative scenario, the weaker negative one. This inverts the standard theory of rational choice (von Neumann and Morgenstern, 1944), on which preferences are primitive inputs. Here preferences are constituted, outputs of the evaluative process, so the theory that models their aggregation is supplied with an account of where they come from.

Chang's (2002) parity cases receive a mechanism on the same basis. Two options are on a par when their total values are close in intensity while constituted by responses directed at different dimensions, so no single comparison resolves the ranking. Parity is real as a description of the pre-choice field. Under compelled selection the modality resolves it, because the agent must orient to act — resolved by choosing, never eliminated by it.

Two commitments need their epistemic weight stated, or the account risks truth by bookkeeping. First, the principle that choice follows the strongest total value would be a tautology if total value were read off the choice. It has content because total value is defined by its constituents, primary response and added associations, each identifiable in principle independently of the choice: by report, by physiological measure, by response to the components in isolation. So defined, the principle is an empirical claim that would fail wherever independently measured total value and compelled choice came apart systematically — though its content is modest, since the composition rule assigns no weights to the constituents, and until it does the principle predicts only that choice tracks the constituents' sum in sign and rough magnitude, not that it tracks a determinate function of them.

The shape of a test is concrete. Take a pair of options differing in a single added association — an owned mug against an identical unowned

one, the endowment paradigm — and measure the components separately: the valuation of the bare good, then the ownership premium elicited independently. The principle predicts that compelled choice tracks the summed measures; the stable premiums of the endowment literature are the nearest existing data of this form (Kahneman, 2011).

Second, the forced-choice diagnostic (§5.2) claims that compelled selection under stated conditions — repeated choice, relevant information present, situational constraint controlled — reveals a stable ordering. The construction-of-preference literature (Lichtenstein & Slovic, 2006) documents reversals tracking elicitation format, obtained outside those conditions. The account predicts they attenuate as the conditions are approached. Evidence that reversals shrink with repetition and market experience exists (Plott, 1996; List, 2003) and is consistent with the prediction without yet testing it under the diagnostic's own conditions, so the prediction is carried as a stake: persistent, undiminished reversal under those conditions would show that nothing is being revealed.

The diagnostic is staked on that convergence, with one confound to control. Having chosen, the agent acquires added value for the option taken, the derived worth of its being hers, which stabilises later choices and can counterfeit a pre-existing order (Chen & Risen, 2010). Convergence must therefore be read across choices insulated from that feedback. A single forced choice remains ambiguous between disclosing an ordering and constituting one; only insulated repetition discriminates.

4.2 Three kinds of choice

Choices classify by the evaluative profile of the options. All-positive: both attract, one more strongly — two songs you enjoy. Mixed: one attracts and one repels, as with the well-paid, soul-destroying job against the poorly paid vocation; this is the most common form in moral life, and typically the most complex. All-negative: both repel, and you choose the lesser evil — the person atop a burning building, choosing between jumping and staying, moved by the difference in intensity of two fears rather than by attraction.

4.3 Reflex, instinct, and intention

Not all behaviour involves this process. Reflexes bypass thought and emotion entirely, since the knee's kick travels a peripheral circuit; no evaluation occurs, so reflexive behaviour falls outside the evaluative domain. Instinctive actions involve emotion without the comparison of alternatives, as the scream at the attacking dog follows the response directly, evaluative but not deliberate. Intentional actions involve the full process: options represented, each evaluated, the responses compared, the strongest acted on. The triad adds a finer grain than Kahneman's binary, whose System 1 spans both reflex and instinct, where the difference between them matters for moral responsibility (Chapter 18).

4.4 Deliberative deadlock and field-enrichment

When two accurate registrations pull opposite ways, there is no scale to read off which is heavier. This section says what deliberation does instead — it widens the field until the field tips — and what is left when it will not tip.

The account must now meet the objection it has seemed to invite since Chapter 14: that a theory on which value is constituted by responses possesses a fact-maker and no scale. Where registrations conflict, as with five patients' suffering against one patient's death, or a friend's present hurt against his future ruin, perfectly accurate empathy delivers opposed pulls of comparable force, so the demand arrives for the correct exchange rate: the fact about which outcome is worse, independent of anyone's weighing.

Why the demand dissolves. An exchange rate valid independently of every weigher is a response-independent evaluative fact, the vacant clause of Chapter 14 (§3.6) in its practical costume, which dissolves as its predecessors dissolved. There is no cosmic scale for the same reason there is no free-standing badness: a ranking no possible evaluative system produces is a ranking composed of nothing. The theorem reaches procedures as well as outcomes, since a procedure every

possible party must register as fair would be the same vacant item one storey up.

The asymmetry this leaves should be stated, because the account leans on it in both directions. Within one evaluator the components of a composite do resolve: they are orientations of a single register, compelled selection is the operation that resolves them, while the constituents are separately measurable (§4.1). Across evaluators no such operation exists, since valence is proprietary to the register that undergoes it (Chapter 14, §3.5) and no register undergoes two parties' pulls at once. The demand for an exchange rate is therefore a demand for a resolution with nothing to perform it.

The demand's strongest form arrives as a consistency argument, which deserves separate audit. Mere addition (Parfit, 1984) and the spectrum arguments descended from it proceed by transitivity and dominance: if each option in a chain is no worse than its neighbour, the endpoints stand ordered, and a theory that refuses the ordering is charged with incoherence.

The reply distinguishes the constraints from what they are asked to do. Transitivity and dominance are orthotic constraints on a comparison scale already in place: once a practice of ranking is adopted, violating them is a defect of the practice, which the account keeps in full force. What they cannot do is institute the scale, because each pairwise step of the spectrum presupposes the commensurability of its adjacent cases, which is the point in dispute. The argument's chain is built from the very comparisons whose availability it was meant to prove.

Where the parties' registrations sustain the pairwise comparisons, the constraints bind inside the practice those registrations constitute. Where the registrations give out, the chain has no links, so the refusal of the endpoint ordering is principled: proprietary registrations do not sum across bearers (Chapter 14, §6.5). The cost of saying so is the one this section has already owned — a field that remains level after full enrichment leaves a choice among tragedies, with the residual work belonging to procedure.

What the demand misdescribes, this chapter's mechanics already describe. A deadlocked choice is not a question awaiting a missing answer. It is a state of the evaluative field, opposed responses in equilibrium, and fields exit equilibrium by the only route they ever had: enrichment. New secondary objects enter — what the choice does to the surviving families, to trust in the institution, to the precedent governing the next hundred cases, to safety, to liberty — each carrying added value of the kinds §3 taxonomised, until the composite tips and motion resumes. Deliberation is not measurement that fails for want of a unit. It is the widening of the field until it ceases to be tied.

The practice confirms the analysis at its terminus. Where enrichment runs out and the options remain level, institutions resort to the lottery, and a lottery is intelligible only on the present account, as one more added value rather than an exemption from the theory. Procedural fairness is a secondary object the parties' registrations can converge on when nothing else differentiates the options, and whether they converge is itself contingent, constituted by their actual responses like everything else in the field.

The practice shows both outcomes, and both are data for the account. Some parties register the coin as the one impartiality left standing. Others register a lottery over lives as an abdication, a refusal to weigh experienced as an insult to the stakes, which is why triage boards and courts have as often rejected sortition as adopted it. A theory of correct weights can explain neither party: not the ones who accept a coin, since beings who possessed the weights would never need one, nor the ones who refuse it, since their refusal registers nothing on that theory but error. On this account both are what they seem — registrations, contestable, belonging to the procedures that follow.

The legitimacy criterion. Enrichment requires a discipline, or the account licenses the manipulator, whose whisper also adds value and also tips fields. The discipline is the standard fixed in Chapter 14 (§2.2): an added value is legitimate insofar as it bottoms out in real registrations. "What this does to the community", pursued downward, terminates in feeling creatures — the families, the future patients, the citizens whose trust is at stake — and is empathy cast over a wider

causal field. The demagogue's installed association is a harder case than that sentence suggests, since a demagogue's whisper usually does bottom out in real registrations — real fear, real loss — and what makes it manipulation is a misreading: the association is keyed to a picture of the affected parties that is false or incomplete, so the criterion must be read as Chapter 21 (§7.6) reads it, as the fittingness of the added value to the registrations it invokes. So read, the whisper that terminates in no affected party and the whisper that terminates in a misdescribed one are classified together, as corruption, by the accuracy-vocabulary that classified the mannequin case.

Weighing, so understood, inherits the account's full error-theory. One weighs wrongly by excluding registrations from the field, or by reading falsely the ones admitted, which is historically how the great moral failures proceeded (Chapter 17, §7.1) — never by miscomputing a sum that was there to compute. Where a fully enriched, fully accurate field remains in equilibrium, the account says what no rival theory says: there is no further fact within that field; the residual work belongs to the operation across fields that the next paragraph names, the deep persuasion of Chapter 21 and the lawmaking machinery of Chapter 22, which are this mechanism at social scale.

Enrichment across evaluators. Across evaluators, enrichment has a form of its own, and the political sequence runs on it. Picture two registrations in conflict as two screens in one hall: the child's suffering on one, the five patients' on the other. There is no third screen on which the two appear weighed against each other. The weighing is in the viewers.

Call a field mutually informed when each party accurately registers the other's registration — when the surgeon weighing five against one has taken in what each of the six undergoes, and each of the six, so far as they can, what the others do. The other party's pull enters this party's field as an empathic secondary object (Chapter 14, §4.2), which is the one object the demand for an exchange rate was looking for in the wrong place.

A verdict on a trade-off is then whatever mutually informed fields settle on. No scale is applied from outside; the settling is reached by deep persuasion, and found nowhere else. The demand for a fact was, underneath, a demand for something shareable, and a shared response supplies that as well as a shared object does. The account stakes one claim on this: mutually informed fields converge more than they diverge. It is the thinnest empirical bet the account makes, marked here as well as in the Register. The contact and deep-canvassing studies Chapter 21 draws on already test the claim, within their limits.

The mechanism has an ancestor. Smith (1759/1976) built moral judgement from mutual sympathy, each party imaginatively taking up the other's feeling and the two adjusting until their sentiments concord, and then idealised the concord into a single impartial spectator whose verdict is the standard. Mutually informed fields are Smith's mechanism without the idealisation: the concord is reached in the parties, with no constructed spectator to reach it for them; where it is not reached, the account records a level field, with no verdict a spectator would have supplied. Harsanyi's (1955) impartial observer, who chooses as though he might occupy any party's position with that party's preferences, is the same move in decision-theoretic dress; the account takes the imagining and declines the sum Harsanyi builds on it, for the reason Chapter 14 (§6.5) gives.

Three outcomes are possible. Where the fields converge on what to do, the convergence is the fact, and nothing further is wanted. Where they do not, they more often converge one level up, on how to decide: the triage rule, the lottery, the vote. Scanlon's individualist restriction, which tests a principle against each person's complaint and never against the sum (1998; Chapter 16, §4.2), is a convergence of this kind stated as a rule. That is the convergence on procedural fairness described above, a registration in its own right.

Where neither level converges — a fully informed field level on the outcome and level on the procedure — the account's residue begins. There is no further fact, only a choice among tragedies, and what remains are Chapter 21's other two modes, enforcement and management. A theory of correct weights would settle none of these

cases either. It would assert that they are settled; the screen in the empty hall shows what that assertion comes to.

Where the argument stands: evaluation has been given its structure (§2), total value has been partitioned into primary and added and the partition derived (§3), and choice has been analysed as compelled selection with deadlock exited by enrichment (§4). What those responses build, across a life, is the evaluative hierarchy.

5 The Evaluative Hierarchy

5.1 Formation

Differing emotional intensities toward different objects generate a hierarchy of value: an ordered ranking of objects by the intensity and orientation of the responses they elicit. Each person's hierarchy is unique, shaped by developmental history, accumulated association, together with the involuntary delivery of emotional content through subconscious processes she does not control. None is static. A trusted friend's betrayal destroys derived value and the friend's position shifts; a career valued for prestige reverses when it is found to harm. The dynamism follows from the composite structure, since any change in the secondary objects can transform the ranking.

5.2 The forced-choice diagnostic

Imagine five faulty devices, each containing a different object — a plant, a snake, a dog, a child, and a precious shirt — each about to activate and destroy its contents. You must decide in which order to rescue them. Different people choose differently, because the choice depends on each object's total value in the rescuer's hierarchy. Primary value is the empathic response the object triggers directly, strongest for the child and weakest for the plant in most human hierarchies. Added value enters where a billionaire's fortune for the shirt may lift it above the plant, or the dog — would anyone place it above the child? The scenario is an intuition pump and is offered as one; what makes the diagnostic a diagnostic is the set of evidential conditions §4.1 fixed —

insulated repetition, controlled feedback — under which a compelled choice discloses an ordering already in place.

Now replace the objects with humans and a dog, then introduce secondary factors. The man is your brother (derived); the woman knows the cure for cancer (consequential); you hold a principle that children are saved first, an orthotic assessment interacting with the empathic evaluation. The hierarchy shifts and complexifies with each importation, so the ranking that emerges is constituted by the totality of responses, each with its own intensity.

The philosophical yield is specific. The hierarchy is a lived structure that manifests in action and cannot be fully articulated in advance of what this chapter calls the forced-choice diagnostic (FCD): the revelation of hierarchical position under the pressure of compelled selection. The evaluator may not know which object she values most until she must choose. The hierarchy is constituted by evaluative responses rather than by cognitive judgements, so the responses emerge under the situation's pressure, sometimes surprising the evaluator herself. The FCD's evidential conditions and its constitutive limit were fixed at §4.1; subsequent chapters reference it by this label. (Gibbard's (2003) plan-expressivism marks the connection to normative theory: if normative judgements are plan-endorsements, the plans are products of the hierarchy the FCD reveals, which is the foundation of normative life.)

5.3 The subconscious hierarchy

A parallel hierarchy operates beneath awareness. Emotional patterns stored in the subconscious (associations with memories, fears, desires, traumas) carry their own evaluative weight and continue to shape reactions, preferences, and choices whether or not they ever surface (Bargh & Chartrand, 1999). A person who inexplicably avoids a kind of social situation may be responding to a buried derived-value connection between that situation and a childhood humiliation. Psychological analysis aims to uncover these layers; the subconscious hierarchy is the involuntarily constituted base on which the conscious one is built.

6 Engagement with Rival Accounts

6.1 Kahneman and dual-process theory

System 1 (fast, automatic) and System 2 (slow, deliberate) map onto the account with significant convergence, to which the account adds two things the framework lacks.

First, it separates two axes the fast/slow contrast runs together: tempo and route (§3.1), which cross. A fast response may be primary, the immediate pull of a colour, or added, the halo an association discharges automatically. A slow one may be primary, attentive savouring of a wine's own character, or added, the deliberate weighing of provenance and consequence. Kahneman's System 1 is therefore not one evaluative kind but two, distinguished by route beneath a shared tempo — the diagnostic the value taxonomy supplies where fast/slow cannot. When System 1 says "attractive" and System 2 says "dangerous", the conflict may be primary-versus-added or added-versus-added, which call for different corrections. Marketing does its work here, installing fast added value that wears the immediacy of a primary response (§7.2).

Second, the analysis identifies what both systems share. Both operate through the evaluative modality, since System 2's reasons are relevant only because they engage it: a reason that triggers no evaluative response is inert, which in the model of Chapter 14 (§A.5) is a field with no body in it. The difference between the systems is breadth of field, not kind.

6.2 Damasio and the somatic marker hypothesis

Damasio's (1994) somatic markers — bodily emotional signals that mark options as desirable before deliberate reasoning begins — converge substantially with the account, which adds the taxonomy his markers lack. A marker says only "this feels good". The account differentiates whether the feeling is constituted by primary, derived, or consequential value, a differentiation that matters clinically: a gambler's marker may respond to the thrill of the bet, which is primary, while

failing to register the financial consequences, which are consequential. The marker is a composite, and a composite can be analysed.

6.3 The rationalist tradition

Decision theory from expected utility (von Neumann and Morgenstern, 1944) to its refinements (Broome, 1991) treats choice as calculation, with emotion as an input given from outside. The account inverts the relationship. The calculation is an orthotic operation on evaluative content the modality has already constituted; without the oriented responses there would be nothing to calculate, no option better or worse than any other.

Broome's consistency requirement illustrates the point. Consistency among preferences is a real orthotic achievement, yet it says nothing about where preferences come from, so a perfectly consistent hierarchy built on distorted responses — a phobic ordering in which the harmless is ranked as threatening — is rational by Broome's criterion and defective by this account's, which is the criterion doing the work when we call it defective.

6.4 Scanlon and the buck-passing account

Scanlon (1998) holds that to be valuable is to have properties that provide reasons — value passes the buck to its grounds. The account agrees that value is no monolithic property, and locates the divergence with a thought experiment: in a world of beings with perfect cognitive faculties and no evaluative modality, do the properties of objects provide reasons? On Scanlon's account they do, waiting to be recognised; on this analysis the world contains no reasons at all, because the modality that constitutes the differences in virtue of which anything could count as a reason is absent. The buck-passing account correctly identifies the features that engage the modality and misidentifies them as self-sufficient sources: they are sources of reasons only because a modality exists that responds to them.

6.5 Bratman and planning agency

Bratman (1987) shows that intentions are not momentary desires but commitments that structure future deliberation; the account supplies what he leaves open, the sustaining mechanism. An intention is a sustained position in the evaluative hierarchy, maintained by the modality's continued positive response to the prospect of completion and negative response to abandonment, losing its structuring power when that orientation dissolves. This explains the familiar discovery that one no longer intends what one thought one intended: the cognitive commitment persists while its evaluative grounding has gone.

6.6 Nussbaum, Elster, and adaptive preferences

Nussbaum (2000) raises the challenge the account must confront directly. If hierarchies are constituted by responses shaped by conditioning, how are real preferences distinguished from adaptive ones, preferences shaped by oppressive conditions to endorse those conditions?

The primary/added distinction supplies the profile. The woman who prefers subordination does not find subordination intrinsically attractive; her preference is constituted by added value — the consequences of non-compliance (punishment, ostracism, insecurity) and the derived associations of the role (approval, belonging). Elster's (1983) sour grapes exhibits the same mechanism, since the grapes' primary pull was never altered, only outweighed by the consequential value of repeated failure. Where the primary response remains intact and merely outweighed, the diagnosis is operationalisable by the removal test: withdraw the externally imposed secondary objects — the punishment, the ostracism, the threatened condemnation — and a preference sustained by added value alone collapses, as the fox's indifference would with the grapes in reach.

The harder cases do not fit the suppression model, and §3.3 has already said why: added value can transform the primary response rather than override it. A person socialised from birth into a subordinate role may have no intact primary response waiting to re-emerge, since

the conditioning has shaped what her responses are — a transformation in the time-indexed sense of §3.2, carrying its history rather than pleading it. For such cases the removal test misfires, because nothing contrary re-emerges, and its error lies in the tacit assumption of a pristine, pre-social primary value constituting the person's real standpoint. The response-constitutive account recognises no such standpoint (Chapter 14, §6.4): there is no vantage outside every evaluative formation from which a person's settled responses can be declared not truly hers.

Two resources support the critique. What grounds the critique of a transformed adaptive preference is therefore not deviation from a natural baseline. It is the two resources that support evaluative criticism throughout the account.

The first is intra-agent consistency. The adapted preference characteristically fails to cohere with the agent's wider commitments, with the empathic response she extends to others in her position, and with the response she forms when the full range of options is before her. The second is convergence, since the direction of her conditioning is one against which human responses are otherwise near-universal. Neither resource pronounces her incorrect against a stance-independent fact, and neither needs to; both license the empathic verdict that her condition is bad and the practical project of altering it.

Deep persuasion is that project's mechanism, now described accurately. It does not restore a buried primary value. It enriches the field (§4.4), introducing new associations and making foreclosed options available, so that a different composite can form — one that coheres with the commitments from which the conditioning had isolated her.

7 Objections and Replies

7.1 The rationality objection

Objection. The account reduces choice to emotional response, leaving no room for rational deliberation.

Reply. It identifies what rational deliberation operates on. Reasons are considerations the cognitive system identifies as relevant, and a reason moves an agent only if it engages the modality — a reason without evaluative engagement is an inert proposition (§6.3, §6.4). Deliberation is the cognitive processing of evaluative content, not the source of it — and, in the deadlocked case, it is the enrichment of the field (§4.4), which is cognition and modality collaborating, not cognition replacing feeling.

7.2 The added value manipulation objection

Objection. If total value is a composite, marketing is value-manufacture: celebrity association, status signalling, and manufactured scarcity inflate a product's added value without touching its primary value.

Reply. The objection identifies a real phenomenon; the account supplies both its diagnosis and its discipline. The inflated value is real as felt, since the consumer really is more attracted, while it is located in the secondary objects, which the severing test exposes: peel the association and watch what dies.

The normative line is the legitimacy criterion of §4.4: added value is legitimate insofar as it bottoms out in real registrations — bottoms out in, not merely produces. The manipulator's whisper produces a registration in the person he moves and terminates in none in the party it invokes, which is the difference. The marketer's installed association, pursued downward, terminates in no affected party's condition, which classifies it with the demagogue's whisper rather than with the surgeon's warning. A consumer equipped with the taxonomy can audit her own composite; a public equipped with it can audit its persuaders (Chapters 21–22).

7.3 The commensurability objection

Objection. The hierarchy assumes all objects rank on a single scale of intensity, but some values seem incommensurable — a friendship and a career cannot be meaningfully compared.

Reply. The hierarchy requires no common metric available to reflection. It requires only that, when a choice is forced, the modality orients — which it does, whether or not the agent can articulate a scale. Unrankable options are Chang's parity, placed at §4.1: where the responses are matched there is no antecedent ranking for the forced choice to disclose, so the choice constitutes a resolution rather than reporting one. The deliberative version of the problem, how conflicting registrations are weighed, is answered by enrichment (§4.4): deadlock is exited by widening the field rather than by discovering the missing unit, and where a fully enriched field remains level there is no further fact to find.

Chang's own later development of the parity cases deserves the comparison, because it offers a rival mechanism for what happens when enrichment has done its work and the field is still level. On her account (Chang, 2017), hard choices are occasions for normative self-creation: the agent can put her will behind one option and thereby confer on herself a reason she did not have before — commitment as the source of new normativity rather than the discovery of old.

The present account keeps the phenomenon and relocates the mechanism. Committing is itself an entry into the field. The act places added value on the chosen side, identity-derived because the option becomes hers, and consequential because abandonment now carries costs that drifting never did; that added value is why commitment binds where velleity does not. What Chang reads as the will manufacturing a reason, the composite reads as the will supplying one more secondary object whose added value tips and then stabilises the total.

The difference matters downstream. Chapter 21's deep persuasion inherits enrichment as its mechanism, and a will that could mint reasons by fiat would reopen the door the manipulation discipline (§4.4) closes. Added value deposited by commitment bottoms out in real registrations, the agent's own, and passes the same audit as every other entry.

7.4 The cultural variation objection

Objection. Hierarchies vary enormously across cultures; if they are constituted by responses, value is completely relativised.

Reply. The account claims a common structure, not a common hierarchy: all hierarchies are constituted by oriented response to primary and secondary objects, differentiated by route into primary, derived, and consequential value. Content varies; structure does not; and the taxonomy is a tool for analysing any hierarchy, not a prescription of what should be valued. Where hierarchies collide morally, the anchoring of Chapter 14 (§6.4) and the machinery of §4.4 apply.

7.5 The reductionism objection

Objection. The account reduces value to emotion, ignoring moral principles, rational commitments, and long-term projects.

Reply. Principles, commitments, and projects contribute to the hierarchy through the modality. “Never harm a child” has force because the modality responds with strong negative orientation to the prospect; “I will finish my degree” because it responds positively to completion. These are forms of added value operating within the hierarchy, whose mechanism the account identifies through which they acquire the importance the objection correctly attributes to them.

7.6 The organic unities objection

Objection. G. E. Moore (1903) held that the value of a whole must never be assumed to equal the sum of its parts, since the consciousness of a beautiful object exceeds the value of the beauty plus the value of the bare consciousness. If organic unities are real, the compositional formula of §3.3 is false in principle, taking the severing test with it: cutting an association changes the whole, so the “components” were never constituents at all.

Reply. The account agrees with Moore about the phenomenon and departs over its mechanism. §3.3 already flags the additive formula as a simplification, since added value can transform rather than combine. What the framework adds is an explanation Moore left brute: the bearer

of value is the oriented response, so the modality can respond to a configuration as such. A composite content can elicit a pull no component elicits in isolation, because the response's object is the Gestalt, which is a content in its own right (Chapter 13, §2.2, on composition as a categorical transition). Organic unities are therefore predicted and mechanised where Moore could only assert them.

The severing test survives because it never presupposed additivity. It is an intervention asking a classificatory question — does this component of the pull survive the cut? — and where the interaction is organic it returns a verdict finer than the case's surface suggests, with the fineness staying inside the partition of §3.5 rather than adding a cell to it.

Configurational value is primary value whose object is the configuration. The Gestalt, as the reply has already said, is a content in its own right, and its contrast with its components is an encounter-delivered relational diff of the kind §3.5's edge-clarification admits. The pull that belongs to the arrangement and to neither relatum alone therefore rides the primary route, remarkable for its composite object rather than for a third channel.

That verdict identifies the cases in which value must be assessed at the level of the whole. It also explains the transformation cases of §3.3 as configurations in which an imported association has been absorbed into the object's registered character, absorption being what conversion means on the time-indexed test (§3.2): the once-imported is now encounter-delivered, and the verdict "primary on the composite" records the completed conversion. What is declined is only Moore's interpretation, since no non-natural property of the whole needs positing.

8 Conclusion

The primary/added distinction is the chapter's most consequential result, because it is the most widely applicable. The great moral conflation operates through it (Chapter 16), the incompleteness of free will is explained by it (Chapter 18), while the possibility of deep

persuasion is grounded in it (Chapter 21). The evaluative hierarchy is the medium in which all of these occur.

The chapter yields four results, all constructed.

The primary/added partition. Total value is the composite of primary value and added value in its two species, individuated by the severing test and derived at §3.5 as an exhaustive partition by logical form. A challenger must reject response-constitution or exhibit a content that neither confines itself to the encounter nor imports beyond it.

The structure of choice. Preferences are constituted, not primitive; choice acknowledges the strongest total response; the forced-choice diagnostic reveals the hierarchy under stated, falsifiable conditions and constitutes resolutions where responses are matched.

The evaluative hierarchy. Unique, dynamic, layered above a subconscious base the agent does not control.

Weighing as field-enrichment. Deliberative deadlock is equilibrium, exited by the entry of new secondary objects. Added value is legitimate insofar as it bottoms out in real registrations, while the lottery is one more added value, its acceptability constituted by the parties' registrations rather than guaranteed by fairness. Across evaluators, enrichment takes the form of mutual registration, so a verdict on a trade-off is what mutually informed fields settle on. Where they converge on the outcome, the convergence is the fact; where only on the procedure, the procedure is; and where on neither, there is no further fact — the residue named at §4.4, where Chapter 21's other modes remain.

If evaluation is structured in this way, with the hierarchy constituted by responses the agent does not control, a question presses. Have the ethical traditions correctly identified the evaluative operation that constitutes morality, or have they systematically confused the tool with the thing it measures? The next chapter argues that they have, and names the confusion.

Grades and Dependencies

1. The three empirical stakes. All four results are *constructed*, on the response-constitutive account (Ch. 14) and the two-axes architecture (Ch. 13). The severing test individuates the partition: a component is added at a time iff severing the association, object held fixed, kills it — with the three admissibility conditions on transformation claims (§3.2). Three commitments are staked empirically: choice tracking independently measured total value (endowment-paradigm form) and FCD convergence under insulated repetition (§4.1) — persistent reversal under the diagnostic's own conditions would falsify — and the convergence of mutually informed fields (§4.4), staked on the contact and deep-canvassing evidence of Chapter 21 (§6.2). The Premise Register carries the chain.

References

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, 54(7), 462–479.

Bratman, M. E. (1987). *Intention, plans, and practical reason*. Harvard University Press.

Broome, J. (1991). *Weighing goods: Equality, uncertainty and time*. Blackwell.

Chang, R. (2002). *Making comparisons count*. Routledge.

Chang, R. (2017). Hard choices. *Journal of the American Philosophical Association*, 3(1), 1–21.

Chen, M. K., & Risen, J. L. (2010). How choice affects and reflects preferences: Revisiting the free-choice paradigm. *Journal of Personality and Social Psychology*, 99(4), 573–594.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. Putnam.

Elster, J. (1983). *Sour grapes: Studies in the subversion of rationality*. Cambridge University Press.

Gibbard, A. (2003). *Thinking how to live*. Harvard University Press.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63(4), 309–321.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Korsgaard, C. M. (1983). Two distinctions in goodness. *Philosophical Review*, 92(2), 169–195.

Lichtenstein, S., & Slovic, P. (Eds.). (2006). *The construction of preference*. Cambridge University Press.

List, J. A. (2003). Does market experience eliminate market anomalies? *Quarterly Journal of Economics, 118*(1), 41–71.

Moore, G. E. (1903). *Principia ethica*. Cambridge University Press.

Nussbaum, M. C. (2000). *Women and human development: The capabilities approach*. Cambridge University Press.

Parfit, D. (1984). *Reasons and persons*. Clarendon Press.

Plott, C. R. (1996). Rational individual behaviour in markets and social choice processes: The discovered preference hypothesis. In K. J. Arrow, E. Colombatto, M. Perlman, & C. Schmidt (Eds.), *The rational foundations of economic behaviour* (pp. 225–250). Macmillan.

Rabinowicz, W., & Rønnow-Rasmussen, T. (2004). The strike of the demon: On fitting pro-attitudes and value. *Ethics*, 114(3), 391–423.

Scanlon, T. M. (1998). *What we owe to each other*. Harvard University Press.

Smith, A. (1759/1976). *The theory of moral sentiments* (D. D. Raphael & A. L. Macfie, Eds.). Clarendon Press.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

Zimmerman, M. J. (2001). *The nature of intrinsic value*. Rowman & Littlefield.

CHAPTER 16

The Great Moral Conflation

Presupposes: the three-cell partition of mental features (Ch. 13, §2.3); the response-constitutive account of value (Ch. 14); the primary/added value taxonomy (Ch. 15).

Establishes: that morality is constitutively empathic; the great moral conflation common to the major ethical traditions; the two-layer analysis and its diagnostic reach; and the descriptive (non-prescriptive) mechanism statement (§2). The three principal results are constructed, with an exposure split note 1 records.

In brief. A person can satisfy a rule and still be morally monstrous; a bureaucrat can execute a procedure flawlessly while facilitating cruelty. Morality and correctness come apart — and this chapter argues the major traditions have systematically confused them. We evaluate a sunset aesthetically, a kind act morally, and a correctly spelled word orthotically — by checking it against a standard — and moral philosophy has persistently treated the third operation as if it were the second. A thermometer is not heat; a moral standard is not morality. The rules are moral *technologies* — indispensable, revisable codifications of empathic insight — and their felt bindingness is borrowed from the responses they codify. The refutation condition is stated openly — a requirement that binds while engaging no feeling and carrying no sanction — and every candidate the traditions supply came back with valence.

---

1 Introduction

A person can satisfy a rule and still be morally monstrous. A bureaucrat can carry out a procedure flawlessly while facilitating cruelty. A clinician can comply with protocol while becoming indifferent to suffering. Such cases expose something basic about moral life: morality and correctness can come apart.

We call a sunset beautiful, a kind act good, and a correctly spelled word right. In each case we evaluate positively, but the operations differ. The sunset is evaluated aesthetically: the evaluative modality responding to sensory input. The kind act is evaluated morally: the modality responding to the emotional consequences of chosen behaviour for others.

The spelling is assessed orthotically: cognitive comparison against a standard. Moral philosophy has persistently conflated the second and third, treating moral evaluation as a species of conformity-checking. Call this the great moral conflation: the systematic treatment of orthotic instruments (principles, rules, standards) as constituting moral evaluation, when they are derivative tools for the orthotic assessment of morality. The charge is structural. It concerns the form an evaluation is given over the base it rests on, not any concealment, and a tradition conflates whether or not its authors ever looked beneath the rule.

The charge has two parts, and they do not depend on the same premises. The first is a misattribution: the traditions credit to a rule, a procedure, or a command a felt bindingness that belongs to the evaluative modality. This needs only the weak premise that bindingness is valence's contribution, so it holds whether empathy is taken to track moral value or to constitute it. The second is the constitutive claim: the instruments codify a value empathy constitutes. This rests on Chapter 14, though on less of it than it seems: the vacancy result (Chapter 14, §3.6) leaves the tracking reading only one contentful object — the affected party's constituted registration — and on that object the conflation verdict survives in substance.

The diagnosis has a sentimentalist ancestry, consolidated here so that the coinage claims only what is new. Hume held that morals derive from sentiment rather than from reason alone (Hume, 1751/1998); Smith (1759/1976, III.4–5) held that the general rules of morality are formed from experience of what the sentiments approve in particular cases, and that regard for the rules, the sense of duty, is derivative, which is the two-layer structure of §3 in its first statement; Slote (2010) grounds moral judgement in empathy directly; Prinz (2007) argues that moral concepts are constituted by emotional dispositions; Nichols (2004) locates the force of moral norms in their connection to affect. The shared ground is real: all five hold that conformity-checking against rules cannot be the whole, or the heart, of moral evaluation.

Three additions mark the delta. The first is the constitution claim in its Chapter 14 form: the evaluative difference is constituted by the oriented response, on pain of the vacancy result. The second is the two-

layer analysis, which relocates the orthotic layer rather than dismissing it — a derivative instrument whose entire authority is banked empathic insight, a placement the sentimentalist tradition gestures at without deriving. The third is the grip-exhaustiveness argument: every mechanism by which a rule can grip an agent is traced to a registration, so the conflation is shown structurally inescapable for rule-first accounts, not merely common. The coinage names the diagnostic pattern; the additions are what the framework contributes to a charge the sentimentalists were first to press. A second lineage, nearer the diagnostic pattern than the metaethics, is taken up in the appendix that completes the survey of the traditions (§A.1).

Section 2 derives the empathic constitution of morality and states what the account asks of its reader. Section 3 develops the two-layer analysis of moral life. Section 4 traces the conflation across the major traditions, the survey completed in an appendix. Section 5 draws consequences; Section 6 answers objections.

2 Morality Is Empathic

These arguments describe a mechanism; they issue no command. Morality, on the account they build, is a natural phenomenon: creatures with valence generate badness-for-themselves, creatures with empathy answer to it, standards crystallise the answering. The mechanism operates in whoever bears it whether or not anything authorises it, as digestion needed no licence before physiology and none after.

A critic may grant the description and ask what nonetheless binds him: what makes the victim’s registration authoritative over his indifference. He is asking for a bindingness that reaches an agent through nothing in the agent; that demand has been met already. It is the response-independence clause of Chapter 14 (§3.6), vacant there and vacant here, returned in the costume of authority. No theory ever possessed what he asks for; §4.4 audits the traditions on this point. What every tradition actually had, the transmission of moral force through responses, this account keeps and names; what it declines is only the pretence that the force ever travelled by another route. The person the

mechanism does not reach is therefore described rather than refuted: inaccurate about the registrations before him (Chapter 14, §2.2), outside the practice keyed to them, answered by that practice with correction and containment.

2.1 The derivation

The constitutive claim follows from results already established. The derivation comes first, so that its cost is visible.

Two premises come from Chapter 14. The first is the response-constitutive account (§3.5): an evaluation is an oriented response, so a verdict with no valence is a classification, and a morality of pure classification is not normative at all. The second is that evaluative responses are individuated by their objects (§§2, 4): what makes an evaluation the evaluation it is, is what it responds to, as what makes a perception chromatic is what it registers.

Two further premises complete the argument. The third is a partition, exhaustive by construction: the object of an evaluation of conduct either includes the felt condition of some party beyond the evaluator, or it does not. The fourth characterises the moral in two conjuncts. Moral evaluation is evaluation proper, whose verdicts count for and against instead of filing; it is the department of evaluation whose subject matter is conduct's bearing on others, since this and nothing else distinguishes a moral objection from an aesthetic or prudential one. The conclusion is then forced: moral evaluation is oriented response whose object includes the felt condition of affected parties — the empathic branch of the partition.

Where the challenger must dig. The fourth premise is where a challenger must dig, and each exit has a stated cost. Deny the first conjunct, and moral facts are inert — a wrongness that counts against nothing (Chapter 14, §3.5). Deny the second conjunct, and one of two positions results. The egoist relocates morality to the self-regarding branch. The position is coherent, but it entails that the wrongness of cruelty has nothing to do with the victim; an evaluation of conduct answerable to no condition but the evaluator's own is, in the book's

taxonomy, prudence under a borrowed name. The formalist makes morality object-neutral (conformity to a principle, whatever the principle concerns), which erases the line between morality and every other conformity practice: etiquette, grammar, the rules of a game. What restores the line, in every formalism actually defended, is the reintroduction of others' weal into the principle's content. The empathic object returns through the side door.

One exit has a tradition behind it and prompts a refinement rather than a bullet. Kant's duties to oneself, against servility and self-destruction, hold moral evaluation answerable to the agent's own condition. The refinement reads the second conjunct with the agent/patient distinction the account already supplies. The moral object must include the felt condition of an affected party; the agent can be among the affected parties, as patient of her own conduct. On this reading, servility and self-destruction are moral failures because they are cruelties that happen to be reflexive. What stays outside morality is only self-regard in the beneficiary sense: prudence proper.

The other-regarding characterisation is grounded in the practice's de facto subject matter, not stipulated. A word about the word is owed here, because Chapter 14 (§2.2) banned "function" from evaluative foundations; the ban stands. What the ban forbids is telic function: purpose-talk, a standard that grades performance and therefore needs an assigner. What fixes the present premise is an observation, which grades nothing and needs no assigner; it fixes the reference of a practice-name, as the modality of Chapter 14 is individuated by its object rather than by a purpose.

The practice the word "morality" names is the one that regulates how conduct bears on those it affects. That is the work its instruments — the prohibitions on cruelty, deception, betrayal — are found to perform when their justification is pursued to the end. The developmental evidence marks the same boundary from below: from early childhood, people distinguish moral from merely conventional violations by reference to harm and welfare, treating the first as authority-independent and the second as not (Turiel, 1983; Smetana, 2006).

The two routes to reference-fixing. Haidt's challenge to that boundary is registered, not dismissed. Across cultures, the moral domain extends to loyalty, authority, and sanctity concerns not obviously reducible to harm (Haidt, 2012). The derivation no longer stakes itself on winning this dispute, because the premise holds on either outcome. If the developmental boundary stands, the reference-fixing goes through as written. If the wide domain stands, its width is what the conflation thesis predicts: traditions codify evaluations into standards (§3.2), the standards are moralised with the evaluations they codify (§4), and a species that had fused the two for millennia would return survey data of Haidt's shape — a moral vocabulary annexing its rulebooks.

On that route, the rival programme's own results draw the core inside the wide domain. The care and fairness concerns are the invariant component, present in every population measured, while the contested triad varies with the tradition and tracks its rulebook — the signature of inherited codification, since the standards differ where the felt condition of the affected party does not. The reference-fixing then lands as an inference to the best explanation — with one circularity conceded, since the explanation invokes the conflation thesis (§4) to account for the periphery, so the second route confirms the thesis only where it is already granted that codified standards get moralised; the first route does not use that, which is why the two are kept apart: the word's stable referent is the invariant core, the periphery is the conflation observed in the field.

The audit of the triad in §4 (each member dividing into empathic core, orthotic husk, prudence, or projection) is applied to a periphery the rival's data has already drawn, so it borrows no boundary from the account it serves. The dependency on the developmental reading thereby drops from a condition on the result to the first of two routes to it, and the moral-foundations findings enter the ledger twice: a challenge to that route, a confirmation of the second's mechanism. The result carries as constructed, by inheritance from Chapter 14. What the derivation moves is the point of challenge — a critic must now reject a named premise, at the cost attached to it.

2.2 Four corroborations

The derived result predicts what moral experience should look like, and four structural features confirm the prediction.

Valence. Orthotic assessment yields judgements of fit: valid or invalid, permitted or impermissible. Moral evaluation yields a felt orientation — approval, indignation, compassion, outrage — directed at the significance of conduct for others. The phenomenology is one of caring; the difference in phenomenal character tracks the difference between the two operations.

Object. Orthotic assessment takes as its object a relation between a candidate and a standard. Moral evaluation takes as its object a state of affairs in another person's emotional life. A system checking an action against the categorical imperative evaluates a conformity relation; a person moved by the suffering the action causes evaluates its impact on another's well-being. No redescription collapses these into one operation.

Dissociation. A person can comply with every moral rule and be morally empty — performing the required actions from habit, fear, or calculation, without empathic engagement — while another violates a rule from moral concern. Orthotic conformity is neither necessary nor sufficient for moral goodness.

Derivation. Moral rules are not self-generating. The prohibition against cruelty crystallises accumulated empathic response to the suffering cruelty causes; the norm against deception derives from the registered harm deception does to trust. When a rule is questioned ("why is this binding?") the justification, pursued deep enough, terminates in an empathic evaluation: because it causes suffering, because it humiliates, because it betrays. A rule that traces back to no empathic evaluation is not a moral rule but a convention.

Three senses of "empathy" must be kept apart, because the word arrives from psychology carrying all of them. Affective resonance is catching another's feeling. Cognitive perspective-taking is representing how things are for another, which a skilled manipulator possesses in

full. The only sense this book employs is the third: the evaluative modality's oriented response to the emotional consequences of conduct for the affected party — a response whose object is the other's condition and whose accuracy is answerable to it (Chapter 14, §§2.2, 6.4). The senses come apart in both directions. Resonance can occur without evaluation, as in contagion; the constitutive evaluation can occur without resonance, since one need not feel the victim's pain to respond to its badness.

The influential case against empathy (Bloom, 2016) targets the first sense; the present account entails its verdict rather than resisting it. A resonance that tracks proximity rather than the affected party's condition fails the accuracy standard of Chapter 14 by the account's own lights. Once the word is dissected, the anti-empathy literature is internal criticism — an audit of the modality's failure modes — and no objection to its constitutive role. The constitutive reading sits with, without resting on, the empirical literatures that trace moral response to empathic capacity: the empathy–altruism experiments (Batson, 2011), the developmental account of empathic internalisation (Hoffman, 2000), together with the primate precursors (de Waal, 1996).

2.3 The aesthete of cruelty

One rival evaluator tests the characterisation where it is strongest to test it. Not the egoist, but the aesthete of cruelty, who meets a cruel act with evaluative delight, finding it elegant, bold, finely executed. His response is a real aesthetic evaluation of behaviour (Chapter 14, §4.1), and he demands to know why the empathic verdict, and not his, is the moral one.

The answer is that the two responses take different objects; the difference is constitutive. The empathic response takes as its object the affected party's emotional reality; it is emotion responding to emotion. The aesthetic response takes the act's form, its shape and execution, with the affected party bracketed. It is emotion responding to what is not emotion, so it carries the counterfactual signature of a response to a surface: it would survive undiminished if the victim felt nothing, or if the same shape were struck with no victim at all (§6.3). The two are

answers to two questions about one act — is it beautiful, and is it right — and morality is the second, by the subject-matter grounding of §2.1. Only the affected-party-including evaluation performs the work the word marks. The aesthete has not produced a competing answer to the moral question; he has answered a different question and misdescribed it.

If he retreats, the retreat is the concession. Granting that the act is wrong while answering that its beauty weighs more with him than its victim, he is no longer disputing the moral verdict; he is confessing a value hierarchy that ranks form above persons (Chapter 15), which is, in the framework's own terms, cruelty itself — the affected party's reality registered and then outranked. Consistent with §2, the account does not claim to compel him. It claims only that he is not right, which he concedes in the very act of outranking the verdict he now grants. The empathic constitution of morality stands derived and corroborated; the next question is where standards fit.

3 Two Layers of Moral Life

3.1 The empathic foundation and the orthotic superstructure

Moral life involves two layers. The first is the empathic foundation: accumulated empathic responses to the emotional consequences of kinds of behaviour, from which a sense of moral good and bad develops. The second is the orthotic superstructure: those responses crystallised into standards, norms, and rules ("don't steal", "don't be cruel") — compressed records of the emotional consequences that kinds of behaviour tend to produce. The rules are codifications of empathic evaluations, and their moral authority is borrowed from what they codify.

A general point underlies this dependence. Authority is a matter of grip: for a requirement to bind an agent — to be more than a proposition she can entertain and lay aside — it must take hold of her, which it can do in only two ways. It may take hold through the evaluative modality, engaging what matters to her so that conforming carries a felt pull. Or it

may take hold through consequence, with a sanction behind it. A requirement that does neither is a classification rather than an obligation: it sorts conduct and leaves the agent untouched. The orthotic standard, considered in isolation, is such a classification. It acquires moral authority only when the empathic evaluation it codifies restores the mattering the bare standard lacks; coercion supplies grip of the other kind, compliance without moral authority.

There is no third source, and the exhaustiveness is not stipulated. The two dimensions of the mental (Chapter 13, §2.3; re-derived at Chapter 14, §3.5) leave only two possibilities for a requirement present to an agent. Either it is felt, engaging the modality, or it is mere content, which of itself moves nothing (Chapter 15, §6.1). Chapter 14 (§A.5) gives the two sources their model: the pull a field, in the physicist's sense, exerts on a body — and the body's absence.

The exhaustiveness still carries a burden, because the rationalist tradition contains explicit candidates for a third source. Parfit (2011) holds that some reasons are fundamental, their authority admitting no further explanation. But an authority that consists in inexplicability, presented to an agent, is indistinguishable from a classification with a full stop after it — unless the counting is felt, at which point the first source has been engaged after all. Scanlon (2014) locates the purchase of reasons in rational self-governance: to see a consideration as a reason is already to be governed by it. But the seeing that governs is the very thing §3.3 dissects. Either it carries the felt must, so the governance is the modality's, or it is a registration that a relation obtains — and a registered relation, like any content, of itself moves nothing. Broome (2013) is the candid case: he concedes that rationality's requirements may not be normative at all, locating the transition from requirement to motivation in the agent's psychology. That is the concession the grip argument predicts the survey should end in.

The division of labour should be fixed, because the argument has two supports of different strengths. The exhaustiveness is carried by the partition: two dimensions of the mental leave two ways a requirement can be present to an agent, so a third source of grip would refute the partition itself, at the results of Chapters 13 and 14 it rests on. The

survey of candidates is the partition's field test, each rationalist proposal checked for where its authority lands. It stands to the derivation as Chapter 2's convergence stands to its dilemma (§2.2, §2.4): corroboration that the derived result is not idle, with a defeat condition of its own. §4.4 states what a defeating candidate would have to be.

3.2 Standards as moral technologies

The orthotic superstructure is a set of moral technologies. These are devices for preserving and extending first-order moral concern in forms more durable, public, and administrable than immediate empathic response. The term is chosen deliberately: a technology's authority derives from how well it serves its purpose, not from its intrinsic properties. A thermometer is not heat; a moral standard is not morality. The reading has a neighbour in Kitcher (2011), for whom ethical practice functions as a social technology remedying recurrent failures of altruism; the present account takes the technological reading while relocating the function, since what the devices serve is empathic evaluation itself rather than cooperation at large, so success is measured against the accuracy standard of Chapter 14 rather than against social stability.

The virtues of the technology explain its indispensability. A prohibition against humiliating treatment compresses first-order empathic insight into a teachable form, coordinates expectations among strangers, protects individuals when no one present is empathically attuned, and grounds criticism and enforcement. These virtues explain why moral technologies are indispensable at scale, and why treating them as derivative locates their authority instead of diminishing them. When the codified evaluations change — when new empathic insight reveals that a standard permitting slavery, subordination, or persecution is inadequate — the technology must be revised. When a protocol conflicts with the empathic evaluation it was designed to serve, the evaluation has moral priority, because the rule's authority was borrowed from it.

Why, then, do philosophers so often invert the layers? Standards are explicit and formulable while empathic engagement resists

regimentation, so theory-building gravitates to the codifiable. Institutions need protocols, which displace the foundation by making conformity the primary criterion. Children are taught rules before they develop the empathic attention the rules encode — though a child who recites rules without grasping why taunting wounds has not yet become morally competent.

3.3 The grip of reason: epistemic normativity

The grip argument invites a companions-in-guilt reply, the sharpest instrument in the rationalist's case. Epistemic requirements — "believe what your evidence supports"; "accept the conclusion of a valid argument whose premises you accept" — appear to bind by neither route. They are not enforced, their authority does not seem to wait on anyone's feelings, and evidence counted in favour of conclusions before there were minds. If epistemic authority is a third source, the exhaustiveness of §3.1 fails, whereupon the moral rationalist will import the same source back into ethics.

The reply distinguishes two things the objection runs together — the two-axes architecture applied to reason itself. The first is correctness: whether a conclusion follows, whether evidence probabilifies. Correctness conditions are structural through and through, relations among contents — fully as response-independent as the objection says. A proof is valid whether or not anyone feels its force. What obtained in the mindless world was correctness's raw material: structure and reliable correlation.

The second thing is bindingness: the ought in "you ought to believe it". Present a valid argument to an agent and ask what, in her, its authority could consist in. If a felt must (the pull toward the conclusion, the discomfort of the abandoned premise), then the authority grips through the modality; the epistemic case confirms the first source rather than escaping it. If she merely classifies the argument as valid while nothing in her answers to the classification, she has a filing verdict, which moves nothing of itself. The felt must of proof, the experience the rationalist cites as reason's own authority, is valence under a rational description, just as Achtung (§4.1) is valence whose object is the moral

law. Reason supplies the structure; valence supplies the grip; the appearance of a third source is the seam between them.

One consequence closes the objection's sharpest form before it is pressed. If the felt must is valence, the crank feels it toward his circle-squaring as the mathematician feels it toward her proof; the difference between them, the objection runs, has dissolved into enthusiasm. It has not, and the account's own fittingness machinery says why (Chapter 14, §6.4). The alethic pull, like every oriented response, is valence coupled to a registration; the registration — that the proof is valid, that the evidence supports — is accuracy-apt on the correctness side this section made response-independent. The mathematician's pull rides an accurate registration. The crank's rides a misregistration, so his state is the alethic analogue of the phobic's fear: real valence, unfitting coupling, criticisable for registering falsely rather than for feeling. The authority of argument was therefore never the pull alone; it is the pull's answerability to the correctness facts its content registers.

That the seam goes unnoticed in the epistemic case has a structural explanation. Orientation toward truth is nearly a constant of cognitive life; an agent wholly indifferent to whether her beliefs are true is barely an agent. A bindingness felt by virtually everyone is easily mistaken for one owed to no one's feeling, as a lamp that is always on is mistaken for daylight. The moral pull, engaged more variably, exposes the seam the epistemic case conceals.

The consequence for this book's own standing follows, which the book accepts; it was stated at §2. The arguments here are valid or invalid on the logical axis, response-independently; their claim on a reader operates through her orientation toward truth. The rationalist will call this a reductio — a philosophy whose arguments bind only the willing. It is the position stated in the open, and less is conceded than the phrase suggests. The unwilling escape the motivation, but no party escapes the fittingness verdict, which is fixed by the correctness half and owes nothing to anyone's willingness. His arguments, too, move only those who care whether they are sound; "failing", pressed for its content, is a conformity verdict unless it comes to something felt.

Where the argument stands: morality has been derived as constitutively empathic (§2), moral life has been split into an empathic foundation and an orthotic superstructure whose authority is borrowed (§§3.1–3.2); the grip argument has closed the last escape by tracing even epistemic bindingness to the two sources (§3.3). The survey that follows tests this structure against the major traditions, one by one.

4 The Conflation Across Traditions

In each major tradition, an orthotic instrument — a tool for checking behaviour against a standard derived from prior empathic evaluation — is presented as constituting moral evaluation itself. The table below states the pattern for each tradition: the instrument it elevates, together with the place where the borrowed force shows through. The survey then argues the two hardest cases in full (§§4.1–4.2) and the tradition nearest the foundation (§4.3), then names the common structure and its refutation condition (§4.4); the remaining traditions and the tradition that anticipated the diagnosis from outside the canon complete the survey in the appendix (§§A.1–A.2):

Tradition	**The instrument**	**Where the borrowed force shows**
Kantianism	Universalisability test	Achtung — a felt pull, credited to reason (§4.1)
Contractualism	Reasonable rejectability	“What it would do to people” drives the rejecting (§4.2)
Divine command	The command	Euthyphro: the wrongness precedes the forbidding (§A.1)

Tradition	The instrument	Where the borrowed force shows
Rule-consequentialism	The rule-set	Rules derived from empathic evaluation of outcomes (§A.1)
Virtue ethics	The exemplar as standard	The phronimos's own perception is the modality at work (§4.3)
Care ethics; Murdoch	—	No relocation needed: force located in the response from the start (§4.4)
Early Confucianism (Mengzi; Wang Yangming)	—	No relocation needed: the sprout of compassion as primary response; ritual as its instrument (§A.2)

Sustained treatment goes to the two hardest cases and to the tradition nearest the foundation; the others are dispatched by the same diagnostic in the appendix (§A.1).

4.1 Kantianism

The categorical imperative, in its first formulation, is a universalisability test. A maxim is compared to a standard; conformity is registered. The operation is structurally orthotic. Kant understood the test as constituting moral evaluation: moral worth attaches to action from duty, not from inclination or sympathy (Kant, 1785/1998, 4:398). Emotion is explicitly excluded from the foundation. This is the conflation in its purest philosophical form.

Kant did not exclude inclination by oversight; his reason must be met before Achtung is relocated, because the parity result of Chapter 14 (§A.2) concedes what that reason feared. A morality founded on feeling, Kant argues (1785/1998, 4:442), cannot be necessary or universal: feeling varies from agent to agent and moment to moment, so what it grounds binds contingently, while the moral law must bind every rational being as such. The account grants the observation and denies the inference. That the pull varies is true; that a law must therefore have a source other than the pull does not follow, because nothing anywhere binds the agent whom nothing moves — not the law, not reason, not the fact — so the necessity Kant wanted is a capacity no foundation possesses, and its absence from the empathic foundation discriminates against nothing (Chapter 14, §A.2). That reply uses the premise that favouring is grip, costed at Chapter 14 (§A.5); a Kantian who refuses it takes the externalist's exit at the cost recorded there. What is universal on the account is the pull's object: the felt condition of the affected party, present whatever the evaluator feels (Chapter 14, §6.4), which is what a universal law was wanted to protect. Kant located the universality in the form of the law because he could find it nowhere in the sentiment; the account locates it in the patient, where the sentiment answers to it. The relocation of Achtung that follows therefore completes this reply.

The strongest Kantian defence is Korsgaard's (1996a). On her reading, the kingdom of ends is the moral core of the system — rational agents construct principles by treating each other as ends — while the universalisability test is a formal expression of this mutual recognition, not its source. The defence brings Kantianism closer to the present view than Kant's own formulation does, which reveals the conflation rather than dissolving it. If the moral content of the imperative derives from the recognition of persons as beings whose well-being matters, the question is what constitutes that recognition. On the present account it is an empathic evaluation; Korsgaard calls it respect for rational nature. The first formulation's test is then an orthotic instrument codifying that recognition, and Kant's subordination of inclination to duty elevates the instrument over the foundation from which it borrows its authority.

Kant's own affective resource confirms the diagnosis. He acknowledges a feeling — Achtung, respect for the moral law (Kant, 1788/1997, 5:73) — but produced by reason and directed at the law, not at the persons the law protects. The account does not deny the feeling; it relocates its source. Achtung is a felt, sub-personally delivered pull, a directional difference (Chapter 14), and the grip argument shows why any binding requirement must be felt in just this way. Kant is right that something binds and right that it exceeds ordinary inclination; he has mislocated its source, reading as a deliverance of reason what is the modality felt under a rational description. The disagreement is not whether the bindingness is real but whence it comes: from the rule, or from the response the rule codifies. His instruction to cultivate the sympathetic feelings as a means to beneficence (Kant, 1797/1996, 6:457) is the inversion at its clearest: the modality enrolled as the rule's instrument, where on the account the rule is the modality's. The same relocation applies to every tradition below; it is set out here because Kant's is the case where the force is most explicitly assigned to reason.

4.2 Contractualism

Scanlon's (1998) contractualism holds that an act is wrong if the principle permitting it could reasonably be rejected by anyone affected. The operation is orthotic in form — act compared to principle, principle tested against a standard — but the content of reasonable rejection is tied to what the act would do to persons, and the "because of what it would do to people" is an empathic evaluation. Contractualism is therefore the most illuminating case: an empathic foundation dressed in orthotic architecture so well tailored that the empathic core is nearly invisible.

The diagnosis must meet the reply Scanlon would make. Reasonable rejection, he holds, turns on the reasons persons have to object to a principle — the burdens it imposes, the interests it sets back, the deception or unfairness it licenses — and reasons are not feelings: whether a complaint would be fair is a matter of what was promised and what was done, decidable by someone who feels nothing. The reply relocates the base by one step and does not change it. A burden is a

burden because of what it does to the person who bears it; an interest is set back only if there is a condition of the person for it to be a setback of; and deception and unfairness wrong someone only where there is someone for whom being deceived or passed over is a condition they would register, under accurate information, as against them (Chapter 14, §§6.4–6.5). Trace any reason for rejection down and it lands on a registered condition of an affected party, or it lands on nothing; the fork Chapter 14 (§3.6) closes for "value" closes for "reasonable" by the same route. Scanlon's test is therefore not orthotic over an orthotic base.

The form does one thing responses cannot do, which earns a credit before it is called a codification: reasonable rejectability is individualist by construction — a principle is tested against each person's complaint, never against the sum of complaints — and that is a substantive verdict on aggregation, one Chapter 15 (§4.4) cannot reach from the responses alone, since registrations do not sum and the account returns no fact about how they should be weighed. Scanlon's form supplies a rule for the one case where the account's field has no further fact; the two-layer analysis says so: the rule is a codification at its base and a decision procedure at its top; the second is work the base does not perform. It is orthotic over the informed registrations of everyone the principle touches, compared across them — which is the mutually informed field of Chapter 15 (§4.4) described as a rule. That is placement and conviction at once: the theory's base is the account's own, and its form is the conflation, whether or not its author looked beneath the rule.

The contractualist who determines that a principle is not reasonably rejectable, then acts on it without empathic engagement, is checking conformity to a test — the conflation in its most sophisticated form. Scanlon's framework deserves the highest compliment the present view can pay: the compliment of being correctly identified. It is the finest orthotic expression of the empathic constitution of morality. A contractualist who accepted the diagnosis would abandon nothing; she would recognise the framework as an exceptionally good moral technology rather than as the constitution of moral evaluation itself.

4.3 Virtue ethics and its neighbours

Virtue ethics requires more care, because it is the tradition most attentive to the foundation, so the charge against it is correspondingly thinner: on the reading defended below, the conflation enters in the tradition's theory-form and pedagogy, and leaves its account of the virtuous agent untouched. Aristotle's phronesis already acknowledges calibrated emotional response as constitutive of moral competence: the virtuous person feels the right emotions, at the right time, toward the right objects (NE II.6, 1106b21–24). On the perceptual reading (McDowell, 1979; Wiggins, 1980), the phronimos does not apply a standard but perceives the situation's moral features through cultivated responsiveness, which may itself be the evaluative modality in its empathic deployment. The conflation enters elsewhere — in the tradition's use of the phronimos as a standard for others. "Act as the virtuous person would" converts the empathic foundation into an orthotic surrogate: well-calibrated, because the exemplar's responses were appropriate, yet operating as a standard rather than as living evaluative engagement.

The neo-Aristotelian variants instantiate the same structure at different depths. Hursthouse's (1999) test — would a person of practical wisdom do this? — is conformity-checking against an encoded character, and her v-rules — act justly, act charitably — are standards by construction, so the exemplar functions as a standard in the theory and not only in its pedagogy. Foot's (2001) natural goodness supplies a standard against which character is assessed. Annas's (2011) skill model comes closer to the foundation while still measuring excellence against a codified conception of good moral perception.

Murdoch (1970) stands apart. Her "attention" — seeing the other clearly, without the distortions of ego and fantasy — is the evaluative modality directed at the other's reality, the tradition's closest analogue to the present account. Care ethics (Noddings, 1984; Held, 2006) comes closest of all. Its insistence that moral response is constituted by attentive engagement with the particular other, not by rule-application, is the empathic foundation identified from within; the internal care-ethical debate over principles is, in the present vocabulary, a debate

about how much orthotic structure the foundation needs, not about whether the foundation is empathic.

4.4 The common structure

Across the traditions the conflation takes one form: a derivative instrument presented as the thing it serves. The instrument may be indispensable and may usually track empathic adequacy; but when instrument and empathic evaluation diverge, the person who has internalised the conflation follows the instrument.

The refutation condition. The diagnosis is guarded against the charge that it cannot lose. What would refute it is a third source of grip (§3.1): a requirement that moves an agent while engaging no evaluative response of hers and standing under no sanction. The traditions were surveyed as candidate producers of that third source, and each reintroduced valence. Kant's Achtung is a felt pull; the contractualist's reasonable rejection is driven by what the principle would do to people; the divine command, pursued past "God said so", terminates in the suffering it forbids (§A.1). A tradition whose bindingness survived the subtraction of both valence and sanction — a pure ought, authoritative, unfelt, unenforced — would break the pattern. The rationalist's response-independent authority is the one serious attempt to exhibit such a thing; §6.4 meets it with an argument that could have failed. On the model of Chapter 14 (§A.5), the condition asks for a pull with no body; the survey's verdict is what the model predicts. A challenger who produces the third source produces a body the model did not count.

The guard's shape needs naming. Chapter 3 (§5) drew it for the partition; the same charge lands here. A test is rigged when its specification is fulfillable and the theory has arranged to pass it. It is empty when the specification is unfulfillable, and an empty test convicts the theory only if the emptiness was stipulated rather than derived. The condition above is unfulfillable if the grip argument is sound — that is what the grip argument says.

But the emptiness is derived from the partition (§3.1), with the survey as its field test over the strongest candidates the traditions

supply; the survey could have ended otherwise. Kant's Achtung could have proved a recognition with no felt aspect. Scanlon's reasonable rejection could have terminated in a fact about principles rather than about people. Broome could have located rationality's requirements outside the agent's psychology instead of conceding them to it. Each was checked; each came back with valence. The emptiness of the refutation condition is therefore a finding about the field, arrived at candidate by candidate — not a specification written so that nothing could meet it.

Seen from the other side, the structure has a positive face. Each tradition reports something real: a felt bindingness, the sense of being required rather than merely classified. That report is no error to be explained away — it is a directional difference of just the kind the grip argument shows any binding requirement must be. What each tradition mistakes is the source: Kant assigns the force to reason, the contractualist to a procedure, the divine-command theorist to the will behind the command, the action-guiding Aristotelian to the exemplar's character. In every case the force is the modality's own, felt under the description the tradition supplies and credited to the instrument that borrowed it. The conflation is what occurs when borrowed authority is taken for an original one. (Care ethics and Murdoch need no relocation: they locate the force in the response from the start; early Confucianism joins them in §A.2.)

The empirical corroboration. The empirical literature corroborates the two-layer structure from below. Haidt's (2001) finding that moral judgement is typically driven by intuitive emotional response, with reasoning as post-hoc rationalisation, is the empathic priority thesis in psychological form; the dumbfounding data it rests on is sorted, on the account's side, at Chapter 14 (§A.9). His moral foundations theory (2012) then raises a challenge the orthotic/empathic distinction resolves rather than suffers. Of the six foundations, care and fairness bottom out in empathic responses to persons; the contested triad divides under the question.

Each member divides under the question. *Authority:* pursue the justification of "respect your elders" to its end and it arrives at the

elder's felt condition (disrespect wounds, and age is vulnerable to it), with the unknowing elder carried by the informed-registration anchor of Chapter 14 (§6.5). Where no registration is reachable even in principle, the honoured ancestor beyond all knowing, what remains is a rule outliving its object — the orthotic husk of a lapsed empathic kernel, and the conflation's purest specimen.

Loyalty divides at a line the tradition's own vocabulary marks. Betrayal is moral because it injures a person who trusted — the care foundation, never distinctively loyalty's. What is distinctive, preference for the group beyond anything trust requires, is an associative preference and no moral matter. Its moralisation is the conflation at work. The survey's own finding that the triad's weight varies across populations while care's does not is, on the registration-correction account of moral progress (§6.7 below), the demotion in progress.

Sanctity divides three ways. Protection of others from offence is empathic, offence being a felt condition of affected parties. Protection of oneself from accusation is prudence. The residue — private desecration, no witness, no offended party — is the agent's own revulsion projected onto the object as a fact about it, the mechanism of Chapter 8 (§3.4), sorted on the account's side with the dumbfounding data above. In the religious case the offended party is the posited God: sanctity stands as empathy toward a party whose existence Chapters 9–12 examine. A traditionalist will reverse the first audit's direction: the disrespect wounds because elders deserve respect, the desert prior and the suffering downstream. The reversal is the fittingness objection in local dress, met at Chapter 14 (§6.4): a desert specified independently of every felt condition is the vacant clause returning.

The width of the surveyed domain is therefore the conflation's footprint — the return a species that had fused standards with evaluations for millennia would give a survey — and §2.1 uses it as the second route to the reference-fixing. The persistent disagreement between moral liberals and moral conservatives offers itself as an illustration: on this analysis it may be in part a disagreement about whether "wrong" means empathically or orthotically wrong, a reading the vocabulary makes available and the chapter stakes no result on,

since whether the parties divide that way is a question for the sociology of moral opinion.

Greene's (2013) dual-process findings map onto the two layers, though the mapping crosses his labels and should be stated with that crossing visible. His automatic emotional responses are the ones that deliver the characteristically deontological verdicts (the refusal to push the man from the footbridge), while his controlled processing delivers the utilitarian ones. On the present reading the automatic response is the empathic foundation firing at a registered harm, and the utilitarian sum is an orthotic device for making registrations commensurable (Chapter 14, §6.5). The deontological intuition therefore sits on the empathic layer and the consequentialist calculation on the orthotic one, which inverts the usual assignment of feeling and rule. The experimental dissociability of the two systems is what the two-layer analysis predicts but could not demonstrate on its own.

5 Consequences

5.1 The dissolution of the deontology/consequentialism opposition

Moral evaluation, on this account, has two components: an object and a response. The object is the chosen behaviour, chosen in the sense Chapter 14 (§5) fixes; the response is the empathic evaluation of that behaviour's emotional consequences for others. The deontologist is right that morality concerns choices — we do not morally evaluate asteroids. The consequentialist is right that the evaluative modality responds to the impact on others' emotional lives. The opposition dissolves because the insights are complementary, deontology capturing the object and consequentialism the response; each tradition goes wrong by treating its structural insight as the whole. The dissolution also answers the worry that the chapter collapses into consequentialism. The empathic evaluation is of chosen behaviour, so an agent who produces good outcomes accidentally is not thereby praiseworthy, while an agent who chooses well but is defeated by luck may be — the deontological

insight preserved inside an account whose currency is emotional consequence.

5.2 Structural diagnoses

Psychopathy. The psychopath's profile is not a malfunctioning empathy but less of the modality, as colour-blindness is less of vision: orthotic capacity intact, aesthetic evaluation often intact, empathic evaluation reduced or absent (Chapter 14, §2.2; Blair, 1995; Nichols, 2004). This explains the otherwise puzzling combination of charm, rule-awareness, and moral emptiness, and why the condition resists orthotic interventions: rules and penalties address a layer whose deficit lies elsewhere. Maibom's (2005) refinement, that the deficit may lie in practical translation rather than detection, is compatible, since the impairment is located in the modality's constitutive function, the production of oriented response to others' states, wherever along the path it falls.

His verdict "nothing bad here", delivered over a registered suffering, is false the ordinary way (Chapter 14, §2.2): he is inaccurate about a fact that obtains in the sufferer. Where he instead applies the rule correctly and is unmoved, he is the rare case Chapter 14 (§A.5) separates from the common one. He knows the words and not the music (Cleckley, 1941/1976; Johns & Quay, 1962), holding the vocabulary of wrongness with none of its content — the self-standing circle in the flesh. The reason against his act is inert for him while the act stays wrong.

The compliant perpetrator. The bureaucrat who follows orders while facilitating systematic harm is a subject whose orthotic layer has displaced the empathic foundation. The assessment says correct while the evaluation that would say cruel has been suppressed, outsourced to the institution, or trained out. The diagnosis explains both the phenomenon Arendt (1963) made famous and its remedy: the restoration of empathic contact with the affected parties the procedure has rendered invisible, rather than better rules, which address the layer that already functions.

Moral burnout. The clinician or carer who becomes indifferent through overexposure exhibits the inverse dissociation: the orthotic layer persists (protocols followed, standards met) while the empathic modality is exhausted. The two-layer analysis locates the failure and the treatment (modality-restoration rather than rule-reinforcement) and explains why conscientiousness and moral emptiness can co-exist in the same practitioner without contradiction. With the diagnoses in place, the account turns to the objections its claims invite.

6 Objections and Replies

6.1 The Kantian objection

Objection. The chapter mischaracterises Kant: respect for persons is a rational attitude, and Achtung already acknowledges an affective dimension.

Reply. The chapter engages Korsgaard's strongest reconstruction (§4.1), and grants Achtung fully — as a feeling directed at the law's authority rather than at the persons the law protects. Morality is constitutively directed at the emotional consequences of conduct for others; a feeling directed at the law is not that, however strongly felt. Whether a requirement binding by rational form alone can grip an agent at all is the deeper question, met at §6.4.

6.2 The scope objection

Objection. Many moral evaluations — of fairness, justice, rights — concern structural properties of social arrangements, not the emotional consequences of individual acts.

Reply. They are empathic at their foundation. Why is unfairness wrong? Because it causes resentment, diminishment, and exclusion. Why do rights matter? Because their violation causes suffering, insecurity, and loss of dignity. Structural properties of arrangements matter morally because they have emotional consequences for those who live under them — a bullet the account bites in the open, since it makes justice instrumental to felt condition; a "just" arrangement with

no emotional consequences for anyone would be just only in the orthotic sense.

6.3 The anti-sentimentalist objection

Objection. Emotional responses can be morally wrong. A racist's disgust at a person of another race is an emotional response to a person; if morality is constitutively empathic, it qualifies as moral evaluation.

Reply. The racist's response is directed at the person's perceived category, not at her emotional states. On the taxonomy this is an aesthetic-evaluative response, emotion responding to a perceived property, rather than an empathic one; the classification carries a counterfactual signature that keeps it from being post hoc. An empathic response tracks its object's emotional states and shifts as they shift. The racist's disgust is notoriously insensitive to the other's actual joy, suffering, or fear; it tracks perceived category alone and persists through any revelation about the person's inner life. That profile — a response to a surface rather than to a felt condition — is what classifies it, and the object-criterion (Chapter 14, §§2, 4) was fixed before this case arrived to test it. The two-layer analysis then supplies the corrective mechanism: the prohibition against discrimination is an orthotic instrument codifying the empathic recognition the distorted response fails to track.

A harder version remains: the persecutor whose satisfaction is directed at the suffering itself. This response does take another's emotional state as its object — empathic in object, cruel in direction — and Chapter 14 (§6.4) has located it. Empathy and cruelty are one modality oriented in opposite directions, so the cruel orientation is not exempted from moral evaluation but placed within it, at the pole the account identifies as bad. The moral fact about the victim's treatment is constituted by her response, not his; his satisfaction constitutes only that her suffering is good for him. The taxonomy does not gerrymander the wrongful response out of the moral domain. It places it in the domain, condemned.

6.4 The rationalist objection

Objection. Moral knowledge is possible through rational reflection alone; emotion is epistemically unreliable, and a foundation for morality must be stable.

Reply. The objection conflates identifying what conforms to moral standards — which reason does, and the chapter never denies — with constituting evaluative differences, which it cannot. Reason can determine that cruelty violates non-maleficence; but why is non-maleficence a moral standard? Because cruelty causes suffering, which the modality registers as bad. The rationalist may insist that a standard's genesis in empathic response is separable from its justification: that once derived, the standard has rational authority of its own. If evaluative difference is response-constituted (Chapter 14), the separation is unavailable. A standard detached from the modality retains its orthotic structure, since it can still check conformity, but it has lost the evaluative difference that made conformity matter. Reason operating on a severed standard supplies neither of the two things by which a requirement takes hold (§3.1). It engages no modality; it brings no sanction. It can tell the agent that her maxim fails a test, and cannot tell her why failing it should weigh with her.

The most resilient rationalism, the non-naturalism of Parfit (2011) and Enoch (2011), does not claim a felt authority. It claims that a requirement can be authoritative apart from whether any agent feels its pull, stands under a sanction, or acknowledges it. The account cannot disprove this and does not pretend to; what it can do is place it. An authority that evades both sources of grip does so only by belonging to no mental life at all, which just is a response-independent evaluative fact — the posit Chapter 14 (§§3.5–3.6) shows vacant on independent grounds.

The disagreement is the response-constitutive thesis contested again under another name; this chapter assumes nothing beyond it. Where that thesis holds, the rationalist's authority binds no agent in particular; where it fails, the present account fails with it — in Chapter 14, not here. What the two finally divide on is whether authority must be able to

bear on the agent it binds, the contrast Williams (1981) drew between internal and external reasons. The premise that it must — that favouring is grip — is modelled and costed at Chapter 14 (§A.5), where refusing it costs the content of "must"; the Premise Register lists it among the contested premises, with what falls and what stands if it is refused.

6.5 The moral anti-realism objection

Objection. The chapter presupposes that empathic evaluation constitutes moral properties. Mackie's (1977) error theory, Joyce's (2001) fictionalism, and Blackburn's (1993) quasi-realism each deny this in a different way.

Reply. The metaphysical work is done in Chapter 14, and since the quarantine of the field (§A.4 there) the replies are inheritances. Mackie and Joyce are met by the commitment result. The objective prescriptivity whose absence the error theory converts into systematic falsity is the vacant clause, and a vacant clause cannot be a truth-condition; there is no error of the kind for moral discourse to be committing. A fictionalism that prescribes pretending what is already true has prescribed assertion with stage directions. What both positions correctly registered, the account keeps: there are no mind-independent moral properties. What makes a moral claim true instead is, in the first instance, the affected party's own registration — observer-independent without being mind-independent — so a whole society's concurring comfort no more makes cruelty good than a roomful of viewers under one lamp makes a wall red.

Blackburn's quasi-realism is met by convergence, stated canonically at Chapter 14 (§A.4). What he constructs is structurally identical to what the account derives, reached at the cost of one extra and idle step; the position is distinguished from expressivism by its ontology and from robust realism by the vacancy result. The chapter's presupposition is accordingly the sole contentful occupant of its field, at the grading Chapter 14's register entry records. The empathic identification defended here keeps its own tier and its own exposure.

6.6 The partiality objection

Objection. Bloom (2016) argues that empathy is a poor guide to moral judgement: biased toward the identifiable, the similar, and the proximate, and innumerate about the statistical. The objection is empirically well-supported.

Reply. The chapter's claim is constitutive — what makes an evaluation moral — not a reliability claim about how well any instance of empathy tracks its object, and both claims are true together. Vision constitutes chromatic difference and is subject to systematic illusion; the illusions show not that vision is the wrong modality but that its detection requires correction. Empathy's parochialism and innumeracy are limits of the implementation, describable as limits because the object outruns them (Chapter 14, §2.2). The correction is not empathy's abandonment but its disciplined extension, which is the orthotic superstructure's purpose. Rules, principles, and institutions compensate for the modality's biases by encoding evaluative insight in impartial, scalable form: the orthotic layer as corrective lens, and, where competing registrations must be brought into one field, as the machinery of enrichment (Chapter 15, §4.4). Bloom's own alternative, "rational compassion", is on this analysis empathic evaluation disciplined by orthotic instruments. The account agrees with the prescription and supplies its finer diagnosis.

6.7 The moral progress objection

Objection. If morality is constituted by empathic response, moral truth changes when responses change. Slavery, it seems, was not wrong when slaveholding societies' responses did not extend to the enslaved; it became wrong as sentiment shifted. A constitution view appears to make moral progress impossible in principle, since any change in the constituting responses is just change, so the abolitionist cannot have been right before the alteration.

Reply. The objection would land on a view that lets the evaluator's response constitute the moral fact. The anchoring of Chapter 14 (§6.4), defended there against the representationalist challenges that would

reopen the gap, is the reason this account is not that view. The moral verdict on slavery is fixed by the affected party's own registration: the enslaved person's suffering, her own modality returning its verdict, fully present in 1820. Nothing moral was missing then. What was missing was uptake. The slaveholding society's empathic faculty failed to register a felt condition that existed, the failure sustained by distance, dehumanising description, in-group partiality, and the orthotic codification of the failure itself into law and custom. The account therefore contains a theory of moral progress rather than a vulnerability to it. Progress is error-correction in the registration of patients; what improves is the reception, not the standard. The expanding circle (Singer, 1981) is on this analysis literally that — an expansion of whose felt condition gets registered, not a change in what counts.

Three consequences answer the objection's three charges. The abolitionist was right in the load-bearing sense: her horror registered what the slaveholder's comfort survived only by missing, and registered it before the surrounding sentiment changed. Being right ahead of one's time is registering ahead of one's time. Reform is answerable to a standard — the patients' actual registrations — rather than to the future's tastes, so a society coarsening into cruelty is regress, not an alternative morality. That is the distinction the relativist cannot draw; this account draws it structurally.

The direction of progress becomes a prediction. Correction should run toward the previously unregistered, circle outward from near to far, and be driven disproportionately by testimony and contact — by whatever carries the patient's registration across the distance that was blocking it. That is what the moral history of the last three centuries, from the slave narrative to the asylum memoir, in fact shows; Hunt (2007) documents the mechanism at historical scale for the eighteenth-century rights revolutions, with the first-person narrative as the carrier of registration. The orthotic layer completes the picture. Rules and institutions do more than compensate for empathy's biases: they bank the corrections, holding each expansion of registration in a form that does not shrink when attention lapses. Progress, though not guaranteed, is ratcheted. Buchanan and Powell (2018) reach a compatible shape

from evolutionary moral psychology: inclusivist expansion is real, contingent on circumstances, and reversible when they lapse, which is what a ratchet that banks corrections without guaranteeing them predicts.

7 Conclusion

The great moral conflation is the inversion the traditions display once evaluation is understood response-constitutively. Each systematically reverses the relationship between morality and the instruments designed to codify it. The traditions that built the most sophisticated instruments — Kant, Scanlon — are the ones where the conflation is most deeply embedded, and for a reason: the instruments are so good that they appear to be the thing itself.

The chapter yields three results, all constructed, with the conditionality its derivation earned.

Morality is constitutively empathic. Derived from four named premises (§2.1), corroborated by valence, object, dissociation, and derivation (§2.2), and grounded in the practice's subject matter rather than stipulation, in the reference-fixing sense §2.1 distinguishes from banned telic function.

The great moral conflation. Each tradition credits to its instrument a bindingness that is the modality's own (the misattribution component, which stands even on a tracking reading of empathy) and codifies a value the affected parties' responses constitute (the component staked on Chapter 14). What would refute the diagnosis is stated (§4.4); the strongest candidate, the rationalist's response-independent authority, is placed at §6.4.

The two-layer analysis. The empathic foundation and the orthotic superstructure of moral technologies, whose dissociations diagnose psychopathy, the compliant perpetrator, and moral burnout, whose ambiguity resolves the moral-foundations debate, and whose banking function explains why moral progress ratchets.

If morality is empathic and orthotic standards are derivative, a further question arises: are those standards revisable, or permanently fixed? A standard that has never been defeated is perfect. But perfect is not infallible — and the next chapter argues that the conflation of one with the other is the epistemological mechanism by which revisable moral technologies are fossilised into unreformable dogma.

Grades and Dependencies

1. Grades and exposure. All three results are *constructed*. That morality is constitutively empathic rests on the response-constitutive account and object-individuation (Ch. 14, §§2, 3.5) and on reference-fixing by either of two routes (§2.1); it falls with Chapter 14's constitution claim, and the derivation's point of challenge is a named premise (§2.1). The great moral conflation rests on the grip argument (§3.1) and the survey of the traditions (§4); the charge splits in exposure (§1): the misattribution component rests only on bindingness being valence's contribution and survives on a tracking reading of empathy, while the constitutive component is staked on Chapter 14. The two-layer analysis rests on §§3.1–3.2, with the banking function derived at §6.7; it falls with the grip argument, and a third source of grip would refute the partition it rests on (Chs. 13–14). The descriptive mechanism statement (§2) is the account's self-description and carries the grade of all three results. The reference-fixing premise has two routes (§2.1) — the developmental boundary, or inference to the best explanation over the moral-foundations data — so the Turiel dependency is one route, not a condition. The Premise Register carries the chain.

2. Instruments received and used. Empathy and cruelty as one capacity, with cold theory of mind, comes from Chapter 13 (Level 5) and is used at the persecutor and psychopathy (§§5.2, 6.3). The reduced-modality psychopath and the accuracy vocabulary come from Chapter 14 (§2.2) and are used at §5.2. The affected-party anchor and fittingness come from Chapter 14 (§6.4) and are used at §3.3's crank and §6.7's progress theory. The informed-registration anchor comes from Chapter 14 (§6.5) and is used at the authority audit (§4.4). The commitment result comes from Chapter 14 (§A.3) and is used at §6.5. The severing counterfactual comes from Chapter 15 (§3.2) and is used at the aesthete (§2.3) and the racist signature (§6.3). Field-enrichment comes from Chapter 15 (§4.4) and is used at §6.6's aggregation role.

Appendix: The Survey Continued

The survey of §4 argued its two hardest cases and the tradition nearest the foundation in the body of the chapter. The remaining traditions are dispatched here by the same diagnostic (§A.1); the survey closes with the lineage that anticipated the diagnosis outside the Western canon (§A.2). No result of the chapter rests on this appendix; the refutation condition of §4.4 counts its cases among the candidates the survey tested; the table that opens §4 already states the pattern for each.

A.1 Divine command, rule-consequentialism, and the Nietzschean precedent

Divine command theory meets the Euthyphro dilemma directly. If cruelty is wrong because God forbids it, morality is pure orthotic assessment, and God could make cruelty obligatory. If God forbids cruelty because it is wrong, the wrongness is prior to the command — and on the present account that prior evaluation is empathic. The command is a codification, not the evaluation.

Rule-consequentialism shows the same structure at one remove. It derives its rules from a first-order evaluation of outcomes — an empathic evaluation, since outcomes are judged better or worse by their emotional impact on those affected — then substitutes conformity to the rules for the evaluation itself. The conflation is located in the substitution.

A last precedent belongs to no tradition in the survey; its proximity is to the diagnosis, not to the account. Nietzsche's *Genealogy* (1887/1998) holds that moral codes are deposits of prior evaluative activity whose origin they afterwards conceal, and that the concealment is what lends them their air of self-evidence: the two-layer structure and the misattribution together, reached by history where this chapter reaches them by derivation. The divergence is over what is deposited; it is not adjudicated here: a genealogist reads the empathic foundation itself as a deposit — the slave's revaluation, in which the registrations of the weak were promoted to the standard by which the strong are

judged — so what this chapter treats as the base the genealogy treats as one more codification with a history. The account's reply is the anchor (Chapter 14, §6.4): the sufferer's registration of her own suffering is the undergoing itself, installed by no one; a genealogy of the standards built on it does not reach it. Whether that reply answers Nietzsche or merely locates the disagreement depends on whether he would grant the anchor, which is the question the exchange comes to. Nietzsche reads the underlying activity as the expression of ranked power, so a codification serving the weak is a falsification whose exposure discredits it. The present account reads the activity as empathic registration of what conduct does to those it affects, so codification is a technology which banks those registrations and stands to be corrected against them (§6.7). The genealogy of a standard accordingly does opposite work in the two accounts: it unmasks an authority in one and locates it in the other.

A.2 The Confucian anticipation

The survey has so far run through the Western canon; the diagnostic pattern has an older statement outside it. Mengzi's case of the child at the well (2A6) is the dissociation argument in its earliest recorded form. Anyone who suddenly sees a child about to fall in feels alarm and compassion, and the response, Mengzi specifies, precedes any wish to win the parents' favour, any concern for reputation, any aversion to the child's cries (Mengzi, 2008; Van Norden, 2007). The specification strips away the calculating and rule-facing motives one by one. What remains is the empathic response bare: the modality registering another's imminent suffering before any standard has been consulted.

The four sprouts (siduan) — compassion, shame, deference, the sense of right and wrong — are then offered as the endowments from which the virtues grow, with the sprout of compassion (ceyin zhi xin) placed first. Benevolence, on Mengzi's account, is that sprout cultivated. Read in the present vocabulary, the sprouts are primary empathic responses, so the doctrine that the first virtue grows from the first of them is the priority claim of §3.1, stated some twenty-three centuries earlier.

Mengzi's account of moral development is likewise the framework's, under the name of extension (tui). The sprout responds within a narrow circle; cultivation consists in carrying the response to cases it has yet to reach — King Xuan's sparing of the ox extended toward the people his policies touch (1A7). Extension is field-enrichment (Chapter 15, §4.4) practised on one's own hierarchy: the cases brought before the agent supply registrations the response then covers; the widening is the registration-correction that §6.7 identifies as moral progress. What the framework adds is the anchor beneath the practice. Mengzi grounds the sprouts' authority in what Heaven confers; the present account grounds it in the affected party's registration (Chapter 14, §6.4). The endowment story can therefore be given up while the priority of the response stands.

Xunzi supplies the other layer. His teaching that human nature requires the artifice of ritual (li) before it yields order — crooked wood needing the press-frame, blunt metal the whetstone — is a theory of the orthotic superstructure: standards as instruments that hold and transmit corrections a raw endowment would lose (Xunzi, 2014). On the two-layer analysis, the celebrated dispute between the two thinkers loosens. Mengzi describes the foundation; Xunzi describes the instrument built on it; the question of whether nature is good becomes the question of which layer a moral psychology begins from. The framework sides with Mengzi on constitution, since the standard borrows its authority from the responses it codifies (§3.2), and with Xunzi on the indispensability of the codified layer, whose banking function §6.7 derives. Where a Confucian line hardens the li into an unrevisable inheritance, the anatomy of Chapter 17 applies as it applies everywhere. Where the li remain answerable to the responses beneath them, the tradition holds the two layers in the order this chapter argues for.

Later Confucian moral psychology carries the pattern forward. Ivanhoe (2002) traces the self-cultivation tradition from Mengzi to Wang Yangming, whose liangzhi — the knowing that answers another's condition without deliberation — is the empathic foundation described from within the practice of cultivating it. The tradition therefore belongs, in the table that opens this section's survey, beside care ethics

and Murdoch among the accounts needing no relocation: the force of moral requirement is located in the response from the start, with ritual and standard treated, at the tradition's best, as the response's instrument.

Its inclusion also answers a question the survey would otherwise leave open — whether the two-layer structure is a finding about one canon. The same division, drawn in the same order, appears in a lineage that developed without contact with the traditions surveyed above. With the survey complete across both canons, the two-layer reading stands as §4.4 states it.

References

Annas, J. (2011). *Intelligent virtue*. Oxford University Press.

Arendt, H. (1963). *Eichmann in Jerusalem: A report on the banality of evil*. Viking Press.

Aristotle. (1984b). Nicomachean ethics. In J. Barnes (Ed.), *The complete works of Aristotle* (Vol. 2). Princeton University Press.

Batson, C. D. (2011). *Altruism in humans*. Oxford University Press.

Blackburn, S. (1993). *Essays in quasi-realism*. Oxford University Press.

Blair, R. J. R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, 57(1), 1–29.

Bloom, P. (2016). *Against empathy: The case for rational compassion*. Ecco.

Broome, J. (2013). *Rationality through reasoning*. Wiley-Blackwell.

Buchanan, A., & Powell, R. (2018). *The evolution of moral progress: A biocultural theory*. Oxford University Press.

Cleckley, H. (1941/1976). *The mask of sanity* (5th ed.). Mosby.

de Waal, F. (1996). *Good natured: The origins of right and wrong in humans and other animals*. Harvard University Press.

Enoch, D. (2011). *Taking morality seriously: A defense of robust realism*. Oxford University Press.

Foot, P. (2001). *Natural goodness*. Clarendon Press.

Greene, J. D. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. Penguin.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Pantheon.

Held, V. (2006). *The ethics of care: Personal, political, and global*. Oxford University Press.

Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge University Press.
Hume, D. (1751/1998). *An enquiry concerning the principles of morals* (T. L. Beauchamp, Ed.). Oxford University Press.
Hunt, L. (2007). *Inventing human rights: A history*. W. W. Norton.
Hursthouse, R. (1999). *On virtue ethics*. Oxford University Press.
Ivanhoe, P. J. (2002). *Ethics in the Confucian tradition: The thought of Mengzi and Wang Yangming* (2nd ed.). Hackett.
Johns, J. H., & Quay, H. C. (1962). The effect of social reward on verbal conditioning in psychopathic and neurotic military offenders. *Journal of Consulting Psychology*, 26(3), 217–220.
Joyce, R. (2001). *The myth of morality*. Cambridge University Press.
Kant, I. (1785/1998). *Groundwork of the metaphysics of morals* (M. Gregor, Trans.). Cambridge University Press.
Kant, I. (1788/1997). *Critique of practical reason* (M. Gregor, Trans.). Cambridge University Press.
Kant, I. (1797/1996). *The metaphysics of morals* (M. Gregor, Trans.). Cambridge University Press.
Kitcher, P. (2011). *The ethical project*. Harvard University Press.
Korsgaard, C. M. (1996a). *Creating the kingdom of ends*. Cambridge University Press.
Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. Penguin.
Maibom, H. L. (2005). Moral unreason: The case of psychopathy. *Mind & Language*, 20(2), 237–257.
McDowell, J. (1979). Virtue and reason. *The Monist*, 62(3), 331–350.
Mengzi. (2008). *Mengzi: With selections from traditional commentaries* (B. W. Van Norden, Trans.). Hackett.
Murdoch, I. (1970). *The sovereignty of good*. Routledge.
Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. Oxford University Press.
Nietzsche, F. (1887/1998). *On the genealogy of morality* (M. Clark & A. J. Swensen, Trans.). Hackett.
Noddings, N. (1984). *Caring: A feminine approach to ethics and moral education*. University of California Press.
Parfit, D. (2011). *On what matters* (Vols. 1–2). Oxford University Press.
Prinz, J. J. (2007). *The emotional construction of morals*. Oxford University Press.
Scanlon, T. M. (1998). *What we owe to each other*. Harvard University Press.
Scanlon, T. M. (2014). *Being realistic about reasons*. Oxford University Press.
Singer, P. (1981). *The expanding circle: Ethics and sociobiology*. Farrar, Straus and Giroux.
Slote, M. (2010). *Moral sentimentalism*. Oxford University Press.

Smetana, J. G. (2006). Social-cognitive domain theory: Consistencies and variations in children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 119–153). Lawrence Erlbaum.

Smith, A. (1759/1976). *The theory of moral sentiments* (D. D. Raphael & A. L. Macfie, Eds.). Clarendon Press.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.

Van Norden, B. W. (2007). *Virtue ethics and consequentialism in early Chinese philosophy*. Cambridge University Press.

Wiggins, D. (1980). Deliberation and practical reason. In A. O. Rorty (Ed.), *Essays on Aristotle's ethics* (pp. 221–240). University of California Press.

Williams, B. (1981). Internal and external reasons. In *Moral luck: Philosophical papers 1973–1980* (pp. 101–113). Cambridge University Press.

Xunzi. (2014). *Xunzi: The complete text* (E. L. Hutton, Trans.). Princeton University Press.

CHAPTER 17

Certainty, Perfection, and Dogma

Presupposes: the self-vindication of difference (Vol. I, Ch. 1); the structural dependence of content on contrast (Vol. I, Ch. 2); the two infinities (Vol. I, Ch. 8); the ground/mind conflation (Vol. I, Ch. 10); the great moral conflation (Ch. 16).

Establishes: the two-dimensional classification of epistemic states; the perfection/infallibility distinction; the four-stage anatomy of dogma.

In brief. A clock that has shown the correct time for a thousand years is *perfect*; it could break in the next second. Perfection says no error has been detected; infallibility, that none is possible — a claim no finite evidence can carry. The subjective half of the confusion: the mathematician who grasps a proof and the dogmatist who clings to a creed feel *identical* conviction, because the feeling tracks emotional pull, never logical status. The objective half: to verify omniscience you would need omniscience — only God can prove God — and the omniscient-yet-simple guarantor classical theism requires conjoins the two infinities Chapter 8 opposed. Four stages take a standard from earned authority to severed foundation — and the book applies the verdict to itself: one self-vindicating claim; everything above it perfect, revisable, graded so.

---

1 Introduction

Chapter 16 argued that the major ethical traditions have systematically confused empathic evaluation with orthotic assessment. They treat conformity to a standard as morality itself, when the standard is a derivative instrument whose authority rests on the empathic evaluations it codifies. But orthotic standards are not static. They are moral technologies, compressed records of prior empathic insight, valuable and revisable. What happens when they are declared irrevisable?

The answer is that the standard is severed from the empathic foundation that gives it authority: the instrument designed to serve morality becomes a substitute for it, so the tool displaces the thing it measures.

This chapter identifies the epistemological mechanism of the severance, which operates in two stages. The first (§§2–3) is the illusion of certainty: the absence of any reliable internal mark separating grounded from ungrounded conviction. The claim holds typically: the metacognition literature finds domain-limited signals of fluency and feeling-of-knowing (Koriat, 1993), so what the classification asserts is that no such signal tracks the logical status of a claim; signals of that kind exist. This allows dogmatic closure to masquerade as rational necessity.

The second (§§4–5) is the conflation of perfection with infallibility: the inference from "this standard has produced no detected errors" to "this standard cannot produce errors". Irrational certainty is the subjective condition that makes the conflation psychologically compelling; the perfection/infallibility gap is the objective condition that makes it epistemologically unjustified. Together they explain why revisable standards are fossilised into unreformable dogma.

The chapter proceeds in seven moves. It develops the two-dimensional classification of epistemic states (§2), diagnoses irrational certainty as the epistemological basis of dogma (§3), and establishes the perfection/infallibility distinction (§4). It then presents the impossible evidence argument and the two-infinities argument against infallibility's metaphysical precondition (§5), develops the four-stage anatomy of dogma (§6), applies the anatomy to the slavery case and three further domains (§7), and addresses objections (§8).

2 The Two-Dimensional Classification of Epistemic States

2.1 The logical axis

For any claim P, a question can be posed: can P's negation be truly held? If it cannot (if denying P generates a logical impossibility, or if the denial instantiates what it denies, so that any holding of the negation falsifies it), then P is logically closed. If it can (if there is any way for the negation to be both held and true), then P is logically open.

The axis must be put in terms of holding rather than of bare coherence, because the two come apart at the case the classification most needs, where the divergence is the framework's own finding. The negation of "difference exists" is coherent: Chapter 3's appendix exhibits a model of it, and Chapter 1 (§6.4) concedes that it holds unheld at a mindless state of pure equality. It is nonetheless unholdable, since every holding of it, in every mode, instantiates what it denies.

An axis defined over coherence would file the book's own foundation on the open side; an axis defined over holding files it where §4.2 independently places it. The self-vindication of difference is logically closed: its denial is coherent and cannot be held truly, which is the occupancy result of Chapter 1 (§6.4) and the Route 2 profile of §4.2 below. The claim that God created the universe is logically open, since its denial is coherent, whatever else may be said about it. The logical axis is binary: closed or open.

2.2 The emotional-pull axis

A methodological note. The emotional-pull axis receives its full theoretical treatment in Chapter 14, and nothing here depends on it. The classification needs only that thoughts carry emotional charges of differing intensity, a feature of epistemic life any reflective person can recognise. Affective epistemology describes it independently (Hookway, 2008; Brady, 2013; Friedman, 2019), and Damasio's (1994) demonstration corroborates it from the neurology: the removal of emotional markers degrades epistemic performance without impairing logical competence. The metaphysics of what a charge is belongs to Chapter 14. The classification is neutral among the answers, while the presupposes line records the dependency it actually carries.

The state the classification most needs to name is knowledge without conviction: evidence without endorsement, the logical axis settled and the emotional-pull axis flat. The one-dimensional analysis cannot characterise it at all. That such a state occurs is available to reflection and needs no clinical warrant.

The ventromedial-damage literature has been read as exhibiting it. On the somatic-marker interpretation of the Iowa gambling task (Bechara et al., 1997), healthy subjects develop a felt orientation that steers them from the disadvantageous decks before they can say why, while patients with ventromedial damage can often articulate which decks are advantageous and go on choosing badly. Whether that dissociation is affective, or instead a matter of what the participants explicitly knew, has been contested since Maia and McClelland (2004), whose exchange is unresolved. The classification takes no side, because nothing here depends on it: what it supplies is the coordinates for the state, not the evidence that it obtains.

For any claim P, two competing thoughts are available to the subject. The first is the thought that P could be different: that the contrast between P and not-P is a live possibility. The second is the thought that P cannot be different: that the contrast is closed. Each thought may carry an emotional charge (a positive or negative valence with a certain intensity) that draws the subject toward endorsement or pushes her away from it. The emotional-pull axis asks: which of the two competing thoughts carries the stronger emotional charge?

How do emotional charges attach to thoughts about possibility and impossibility? Through association and investment. A subject who has organised significant aspects of her life around a belief — relationships, sacrifices, identity, community — has made an emotional investment in its truth. The thought that P could be wrong threatens the investment; the thought that P cannot be wrong secures it. The greater the investment, the stronger the pull toward closure and the weaker the pull toward openness.

Smithies (2019) argues that phenomenal consciousness plays an irreducible epistemic role, the felt character of experience partly constituting its epistemic standing for the subject. Whether or not that stronger claim holds, the felt charge is at least causally significant, since it is what makes irrational certainty so resistant to correction by evidence alone. It explains the resistance without conferring rational standing: the rationality of a certainty is fixed on the logical axis (§3), not by the intensity with which it is felt.

One further role of the emotional-pull axis is discharged elsewhere but anchored here. The felt necessity that accompanies grasped proof, the phenomenon Kant read as reason's own authority, is the pull axis engaged by content whose denial is incoherent: valence responding to logical structure. Chapter 16 (§3.3) develops the point into an account of epistemic bindingness generally, for which the classification supplies the machinery. The logical axis carries the response-independent correctness conditions; the emotional-pull axis carries everything that ever gripped an agent about them.

2.3 The four states

Crossing the logical axis with the emotional-pull axis yields four epistemic states that the standard one-dimensional analysis does not distinguish:

	Pull toward *closure*	**Pull toward *openness***
Logically closed	Rational certainty	Irrational uncertainty
Logically open	Irrational certainty	Rational uncertainty

Rational certainty. The claim is logically closed, its denial one that cannot be truly held, while the thought of impossibility carries the stronger emotional charge. Logic and emotional pull align. The paradigm is a mathematician who understands a proof and feels its necessity. The self-vindication of difference is the paradigm case: no holding of the denial is a true holding of it, so the subject who grasps this is drawn toward the certainty that difference exists.

Irrational certainty. The claim is logically open, its denial one that can be truly held, but the thought of impossibility carries the stronger emotional charge. The subject feels that the alternative is unthinkable, despite its being perfectly coherent, because the emotional charge attached to closure vastly exceeds the charge attached to openness. The closure may be felt as impossibility, or as the alternative's being beneath evidence — the second form, which §3 separates from the first,

is the commoner. The paradigm: a dogmatist who cannot entertain the possibility that her belief might be wrong, not because she has evidence of its necessity but because the thought of its being wrong carries no emotional pull.

Rational uncertainty. The claim is logically open, while the thought of possibility carries appropriate pull. The subject recognises that the claim might be wrong, feels the pull of openness, and holds the belief with the tentativeness its logical status warrants. This is the healthy epistemic default for contingent claims. The label names the logical status and no degree of confidence: a contingent claim held with high confidence, its alternative kept improbable and never impossible, is rational uncertainty in the classification's sense, which is the state most well-founded empirical conviction occupies.

Irrational uncertainty. The claim is logically closed, but the thought of possibility carries the stronger charge — the emotional pull exceeds the intensity attached to its impossibility. The paradigm: a student who follows every step of a valid proof, agrees that each inference is correct, accepts that the conclusion follows, and yet cannot bring herself to believe it. Clinical anxiety exhibits the same structure in a different domain: the anxious subject knows that the feared outcome is extremely unlikely but cannot stop feeling that it is imminent.

2.4 Engagement with the literature

Four bodies of literature reinforce the classification's independent credentials. Williamson's (2000) knowledge-first epistemology correctly analyses rational certainty without addressing conviction disconnected from epistemic position, and the emotional-pull axis supplies what his framework leaves out. The epistemology of disagreement (Christensen, 2007; Elga, 2007; Kelly, 2010) operates on the logical axis; the classification reveals that some persistent disagreements are evaluation-disputes sustained on the emotional-pull axis, requiring the transformation Chapter 21 calls deep persuasion rather than evidential reassessment.

De Sousa's (1987) emotion-as-salience and Brady's (2013) attention-based account independently confirm the axis as a real dimension of epistemic life. Stanley's (2005) interest-relative epistemology converges from another side, since what is at stake bears on whether a subject counts as knowing. That places the axis within epistemic life without making it a criterion of rationality, because it marks what matters to the subject rather than whether her certainty is warranted.

Sosa (2007) develops a virtue-epistemological account in which knowledge is apt belief: belief that is accurate (true), adroit (produced by a competent faculty), and apt (accurate because adroit). The two-dimensional classification maps onto Sosa's framework. Rational certainty corresponds to apt belief with appropriate emotional endorsement; irrational certainty corresponds to belief that is emotionally endorsed without being apt. What the classification adds is the explanation of why inapt belief can feel identical to apt belief: the emotional-pull axis does not track the logical axis, so the subject has no introspective access to the divergence.

3 The Illusion of Certainty and the Epistemological Basis of Dogma

The most consequential feature of the four-state classification is that rational certainty and irrational certainty are, as a rule, phenomenologically indistinguishable from the inside. The classification's yield follows: since the "closed" cell holds only logical truths and self-vindicating claims, every empirical certainty is "open," and the diagnosis of irrational certainty is made on the logical axis together with an evidential judgement; the pull axis explains the persistence of the misjudgement. In both states, the thought that something cannot be different carries the stronger emotional charge. The subject in each state experiences the same phenomenal profile: conviction, closure, the felt necessity of the belief. The mathematician who grasps a proof and the dogmatist who clings to a creed feel identical certainty. The difference lies outside the feeling. The

mathematician's conviction is grounded, since the opposite of the proved theorem is contradictory, while the dogmatist's is not, since the opposite of the creed is coherent. Both experience the thought of impossibility carrying the stronger charge, so from the inside the two states are identical.

The claim is structural in one respect and typical in another. Structurally: what would distinguish the states from within is a felt mark of grounding, which the classification leaves nothing to carry such a mark: the felt charge is the same species of state in both cases, and an axis cannot register differences in what lies behind it. The calibration literature corroborates the structure without being needed for it, since extreme confidence dissociates from accuracy across domains (Fischhoff, Slovic, & Lichtenstein, 1977) — which is what one expects if the feeling tracks the charge rather than the grounding.

This is why dogmatic conviction is so resistant to correction. The feeling of certainty is a reliable guide only to which competing thought carries the stronger emotional charge, never to the claim's logical status. Cassam's (2019) epistemic vices — closed-mindedness, dogmatism, intellectual arrogance — receive their mechanism here: closed-mindedness is the disposition toward irrational certainty, the disposition for the emotional pull toward closure to overwhelm the logical openness of contingent claims.

Irrational certainty is the epistemological basis of dogmatic conviction. The dogmatist treats a logically contingent claim as though it were logically necessary: she does not merely believe strongly; she feels the impossibility of the alternative despite its being perfectly coherent. The mechanism can be stated in one sentence: the emotional response attached to closure does not refute the logical openness of the claim; it renders the openness experientially inaccessible. The doubt-thought is not entertained and dismissed; it is not entertained at all, because the competing emotional response is too intense for the alternative to gain a hearing.

If rational and irrational certainty are phenomenologically indistinguishable, how can they be distinguished? The answer is

external, never the strength of the feeling. Where the claim's negation is contradictory, the certainty is rational, since it tracks a real impossibility. Where the negation is coherent, the certainty is not thereby irrational: a coherent alternative may be held as merely improbable, and a certainty that holds it so is perfectly rational.

When certainty becomes irrational. Certainty becomes irrational when it treats a coherent alternative as impossible, or as closed to the evidence that could favour it — when what is logically open is judged closed, or held beyond the reach of revision. The fault is a misjudgement of the claim's status, modal in the first case and evidential in the second, rather than an excess of conviction. The felt charge is the mechanism that installs and defends the misjudgement, not the criterion that condemns it.

The second form is the commoner, the one the anatomy of §6 tracks. A dogmatist seldom declares her creed's denial contradictory; she declines to let anything count against it. This is the diagnostic the theological sequence (Chapters 9–12) applied repeatedly: where the denial of a doctrine is coherent, a certainty that the denial is impossible has mistaken the emotional-pull axis for the logical one.

This verdict belongs to reflective certainty, the certainty a subject has arrived at by questioning a claim and weighing its denial. Sosa's (2007) contrast between animal and reflective knowledge marks the boundary. A certainty held before any such questioning is pre-reflective, the unexamined default with which most belief begins, neither rational nor irrational but simply not yet graded.

Wittgenstein's (1969) hinge propositions — that one has two hands, that the world existed before one's birth — occupy this pre-reflective ground. Their denial is coherent, yet to hold them fast is no error, because the believer has not judged the denial impossible so much as declined to put it in question at all. A hinge hardens into a dogma only at the reflective stage, when the alternative has been raised and is then refused a hearing, judged no real possibility when it demonstrably is one.

The contemporary hinge literature has built an epistemology on this ground, beside which the classification should be set. Pritchard (2016) argues that hinge commitments are rationally groundless without being irrational, since the fixed points that make rational evaluation possible cannot themselves be rationally evaluated. Coliva (2015) extends the notion of rationality so that accepting the hinges counts as constitutive of it. Both are describing, in the epistemology of justification, what the two-dimensional classification locates in its own coordinates: a hinge is a certainty of maximal pull whose logical-axis status has never been put in question, held fast because inquiry needs it held, and neither rational nor irrational for the reason the pre-reflective category supplies, since grading begins only where questioning has.

What the classification adds is the diagnosis of the transition the hinge theorists leave undescribed. A hinge examined and retained on its merits becomes reflective certainty and earns whatever the logical axis assigns it. A hinge examined and then insulated — the alternative raised and refused a hearing — has become dogma by the four-stage route of §6. The difference between Wittgenstein's river-bed and a hardened orthodoxy is the difference between the unexamined and the examination foreclosed.

The illusion of certainty is the subjective condition that makes the conflation of perfection with infallibility psychologically compelling. A standard that feels infallible — that carries the phenomenal profile of impossibility-of-error — feels that way because of the emotional-pull axis, not because of the logical axis. The next section shows that the objective condition is equally unjustified: no finite body of evidence can establish the universal claim that infallibility requires.

4 Perfection and Infallibility: The Structural Gap

4.1 Correctness and perfection

Correctness is the epistemological expression of the orthotic operation. To be correct means to conform to a given standard, as a word's spelling is checked against the dictionary. The operation has

three features. It is standard-relative, since the same word may be correct in one dictionary and incorrect in another. It is evidence-based, since alignment is checked by examining the output against the standard. And it is revisable, since the standard itself may change, and verdicts of correctness change with it.

Perfection is the limiting case of correctness: complete alignment of the output with the standard, with no detected deviation. Two features of perfection are essential for what follows. First, perfection is evidence-based — a discovery. One checks the gymnast's performance against the rubric, the proof against the rules of inference, the text against the standard. Perfection is falsifiable (a single detected error overturns the verdict), replicable (others can check the alignment), and public (the standard and the evidence are in principle available to all). Second, perfection is temporal — an assessment made at a given moment, on the basis of evidence available at that moment. A clock that shows the correct time right now is perfect right now. It may break tomorrow. A moral code that has withstood every challenge for centuries is perfect relative to the challenges it has faced. It may face a new challenge tomorrow.

4.2 The gap

Infallibility is a categorically stronger claim than perfection. Where perfection says "no error has been detected," infallibility says "no error is possible." The question is: can perfection ever entail infallibility?

No. The argument has a determinate form. Perfection is based on a comparison between an output and a standard using a finite body of evidence. Infallibility requires that this comparison would yield the same result under all possible evidence, including evidence not yet available. But no finite body of evidence can establish a claim about all possible evidence, because the space of possible evidence is open-ended. The transition from perfection to infallibility is a leap off the epistemological chain, from evidence-based assessment to evidence-transcending assertion.

The chess player who has won every game is perfect; she could lose her next game. The clock that has shown the correct time for a thousand years is perfect; it could break in the next second. The moral code that has produced no detected errors over centuries is perfect; a new moral insight could reveal an error tomorrow. In each case, perfection is a fact about the assessed body of evidence, while infallibility would be a claim about all possible evidence, which no assessed body can establish however large. In the language of epistemological defeaters (Pollock, 1986), perfection is always defeasible, since new evidence can undercut the assessed alignment, while infallibility declares itself undefeatable by fiat.

4.3 The relationship to certainty

The two-dimensional classification (§2) provides the diagnostic. Perfection warrants rational uncertainty: the standard has survived every test, but the thought that it could fail a future one is coherent. Infallibility claims the status of rational certainty — the standard cannot fail — while for any empirical standard the denial is coherent. The infallibility claim is therefore irrational certainty: the emotional pull toward closure has overwhelmed the logical openness of the claim.

There are two structurally distinct routes to a claim's being always perfect, to infallibility, and only one of them is available.

Route 1: Omniscient guarantee. A claim is infallible because its source knows everything and cannot err. This is what scriptural inerrancy, institutional infallibility, and moral-code infallibility claim. Section 5 shows this route is structurally unavailable.

Route 2: Logical self-vindication. A claim is infallible because its denial is self-defeating, since any context in which the claim is made is necessarily a context in which the claim is true. The occurrence premise of Chapter 1 (§5.3) is infallible in this way: to deny that anything happens is to make something happen, so the making of the denial instantiates what the premise asserts. The foundational result built on it with the composition premise — that at least one difference obtains — inherits the same standing, since the denial of difference is an act with

two components, and two components of a real act are a real difference. These claims are not merely perfect, with no error detected so far; they are always perfect, because the conditions under which they could be false are the conditions under which they could not be made.

The present book's own foundational result, that diversity exists, is infallible by Route 2. The claims derived from it — the structural analyses of grounding, time, consciousness, value, together with the political consequences that follow — are perfect: the best available given the analysis's premises and inferences. They are not infallible by either route, because they depend on premises that could in principle be challenged. The book practises what it preaches: its foundation is logically secure, its superstructure revisable.

The correspondence with the book's own confidence register should be made explicit, because the Register is this chapter's distinction institutionalised. Route 2 is the Register's self-vindicating tier, the only tier that is infallible, and infallible for the Route 2 reason: the denial of a self-vindicating claim instantiates it, so the conditions under which it could be false are conditions under which it could not be made.

The tier holds two premises and the one result built from them; the narrowness should be read as this chapter's own doctrine applied to the book: the occurrence premise, the composition premise, and the result that at least one difference obtains (Chapter 1, §5.3). The vindication law that names the mechanism, together with the exhibitive closure that deploys it, both run through a premise graded established, and a claim is infallible only where nothing it depends on is revisable.

Established, constructed, and proposed results are the three tiers beneath it, as the General Introduction (§5) defines them. Established and constructed results are species of perfection, the best available on the evidence and inferences in hand, falsifiable, replicable, and revisable, differing from each other in how much would have to go wrong and from the self-vindicating tier in kind. A proposed result claims the coherence of an option, sometimes under conditions the grading names, and no more, so it stands furthest from Route 2 of anything the Register grades.

The register is therefore not bookkeeping appended to the argument. It is this chapter's epistemology applied to the book itself: every grading locates a result on the perfection/infallibility line, and no grading above the self-vindicating tier claims what Route 1 cannot deliver — which is why the framework, unlike the systems §6 anatomises, has no Stage 2 available to it.

Where the argument stands: the four epistemic states are classified (§2), irrational certainty is identified as the epistemological basis of dogma (§3), and perfection has been separated from infallibility (§4). What remains is to show that infallibility's precondition cannot be met, and then to anatomise what happens when it is claimed anyway.

5 The Impossible Evidence and the Two-Infinities Argument

5.1 The impossible evidence

The claim that a finite text is infallible is, implicitly, a claim about omniscience. The chain of assumptions: the text is infallible because its author cannot err; the author cannot err because the author is omniscient; the author is omniscient because the author is God. Each link is asserted; none is demonstrated.

But could the chain be demonstrated? Imagine you meet a being who asserts that it knows everything. You might ask it a question — but you can only verify the answer if you already know the correct response. You could ask a million questions, and every correct answer would establish only that the being has answered correctly so far: perfection, not infallibility. To verify omniscience you would need to check the being's knowledge against every possible truth — which requires that you yourself possess every possible truth. Only omniscience can verify omniscience. Only God can prove God.

Any prophet claiming divine authorship of a text faces the same limitation. His own knowledge is finite, so he cannot verify the infinite extent of God's. He may be convinced, since the experience of revelation may carry overwhelming emotional force, but the conviction

is on the emotional-pull axis rather than the logical one: his report is evidence that he had an experience, not that the experience was produced by an omniscient being. The belief that a text is divinely authored, and therefore infallible, remains an unsupported opinion even if the text appears flawless at present, because flawlessness is perfection, and perfection does not entail infallibility.

The argument's core has a lineage that should be owned. Locke's rule that assent to testimony can carry no more certainty than the witness's own knowledge possessed (Locke, 1689/1975, IV.xvi), and Hume's audit of testimony to the marvellous (Hume, 1748/1999, §10), both operate here. What the present treatment adds is the conjunction with the perfection/infallibility distinction and the structural argument of §5.2.

5.2 The two-infinities argument

The epistemological argument of §5.1 shows that infallibility cannot be verified by any finite mind. The structural argument goes deeper: infallibility requires a metaphysical precondition that is not merely unverifiable but structurally incoherent.

To guarantee that a text contains no errors under all possible evidence requires knowledge of all possible evidence. That requires omniscience, which requires a mind of infinite internal distinction: a mind that registers every possible truth as a distinct item of knowledge, distinguishing each truth from every alternative. This is infinite diversity (Chapter 8), the endless extension of internal structure.

But the traditions that claim infallibility, the Abrahamic traditions in their classical formulations, also claim that God is absolutely simple: a being with no internal distinction, no composition of parts, no complexity. This is infinite equality (Chapter 8), the complete dissolution of all structure. The claim of infallibility therefore requires that its guarantor simultaneously possess infinite diversity, omniscience to know all truths, and infinite equality, divine simplicity to be the God classical theism identifies.

Chapter 8 established that these two infinities are structurally opposite conditions. Chapter 10 established that the conjunction of divine simplicity with intelligence is incoherent, and closed the evasion the omniscient guarantor most needs here: that God knows all truths in one simple, non-discursive, non-representational act, so that omniscience demands no internal multiplicity. That reply is stated at Chapter 10 (§§3.1–3.2, §5.2) and is defeated upstream twice rather than merely relocated; the present argument collects it rather than re-fighting it.

The containment fork (Chapter 10, §3.2) makes the point a theorem. What is known is either a determinate content — a knowing of every truth rather than of nothing, and so a distinction in the knower — or nothing truth-specific at all, in which case "knows all truths" has been emptied.

The content-level formulation of the truthmaker argument (Chapter 10, §4.2) closes the last exit on the fine-grained individuation of necessary contents that Chapter 2 (§2.4) licenses. "Knows P" and "knows Q" are distinct contents even where both are necessary, the exclusion operating at the level of content-individuation, contrastive whether or not the knowing is composed, sequential, or representational. The formulation carries that grain condition, recorded at its source, and a reader on the coarse grain retains the containment fork, which suffices for the theorem.

The intra-attribute distinctions required for any cognitive state to be a cognitive state survive the collapse of every inter-attribute distinction (Chapter 10, §4) and every trace of succession into one act, surviving it by derivation rather than by assertion. The chain of assumptions that supports infallibility terminates in a structural impossibility.

This is the chapter's distinctive contribution beyond standard fallibilism. Peirce (1877), Popper (1963), Kuhn (1962), and Quine (1951) argue that infallibility lacks evidential support: that our evidence is always finite and therefore cannot establish a universal claim. The present analysis agrees, but adds: infallibility requires a metaphysical precondition, the conjunction of the two infinities, that is not merely

undemonstrated but structurally unavailable. The infallibility claim rests on incoherent foundations. This refutation is scoped accordingly: it bites only the guarantor that classical theism conceives — omniscient and absolutely simple at once. A theism that relinquishes divine simplicity, as personalist, open, and process theologies do, escapes the structural argument, though not the evidential one above, which by itself suffices to defeat the claim that any finite evidence base attains infallibility.

Plantinga's (2000) reformed epistemology deserves particular attention, because it offers the most philosophically developed attempt to bypass the evidence requirement altogether. Plantinga argues that belief in the great truths of the Christian faith can be "properly basic" (warranted without inferential support) if produced by cognitive faculties functioning properly according to a design plan aimed at truth. The structural argument does not reach him, nor should it be made to: Plantinga rejects divine simplicity (1980), so the designer his design plan requires is a being of infinite diversity, and the conjunction the two-infinities argument targets is not one he holds. What reaches him is the evidential argument, on his own terms. Warrant, on Plantinga's account, is defeasible: a properly functioning faculty can be overridden by defeaters, the design plan can malfunction, and no feature of proper basicality upgrades a warranted belief in a text's divine origin into a guarantee that the text contains no error under all possible evidence. The bypass delivers warrant, which is a species of perfection, and not infallibility, which is all §5.1 needs. The reformed epistemologist's route reaches the same station as every other: well-grounded belief, revisable in principle, declared unrevisable only by the conflation this chapter names.

The convergence objection deserves a brief response: multiple independent lines of evidence converging on one conclusion may seem to bridge the gap. Convergence does strengthen a claim, making it more securely perfect. But multiple finite lines remain finite. Convergence increases the degree of perfection without eliminating the gap between "no error detected across multiple independent checks" and "no error possible", so the objection mistakes very high-quality perfection for infallibility, which is the conflation this chapter identifies.

6 The Four-Stage Anatomy of Dogma

The conflation of perfection with infallibility generates a specific pathology. The inference runs from "this standard has produced no detected errors" to "this standard cannot produce errors", so the pathology operates through a four-stage mechanism identical across scripture, institutional authority, moral codes, and scientific consensus.

Stage 1: Earned authority. A standard is established that aligns with the available evidence: a text with no detected errors, an institution whose judgements have been consistently adequate, a moral code that has served well, a scientific theory that has survived every test. The standard has earned its authority through demonstrated alignment. This is perfection.

Stage 2: Extended permanence. The standard's track record is extended into a claim of permanent adequacy. The inference — sometimes explicit, sometimes tacit — is that because the standard has produced no errors, it cannot produce errors. This is the epistemological break: the transition from a finite body of evidence to a universal claim about all possible evidence.

Stage 3: Foreclosed revision. The claim of infallibility forecloses revision. The standard is no longer subject to challenge, because any apparent challenge is reinterpreted as the challenger's error. Anomalies are explained away, dissenters are marginalised, while the evaluative engagement that produced the standard is replaced by conformity to it. The stage has a valid-looking argument behind it, the dogmatism paradox Kripke (2011) found and Harman (1973, pp. 148–149) first reported: if one knows that P, any evidence against P is misleading, so one may disregard it in advance. The resolution is Harman's, which the anatomy makes its own: acquiring the evidence changes what one knows, so a knowledge one holds at a time licenses no policy of ignoring what has yet to arrive. Foreclosure claims that licence, which is why even a standard known true at Stage 1 has no title to Stage 3.

Stage 4: Severed foundation. The standard is severed from the evaluative foundation that gave it authority. The moral code generated by empathic insight is no longer answerable to empathic insight. The

scientific theory generated by empirical inquiry is no longer answerable to empirical inquiry. The standard has become self-authorising — a closed loop that derives its authority from itself.

The mechanism mimics the structure of self-vindicating certainty (Chapter 1). Infallibility claims the status of self-vindication for claims that do not possess it, for which the classification supplies the diagnosis: the felt impossibility of error is on the emotional-pull axis, not the logical one. The four stages are structurally identical across domains, since any standard whose authority is derivative becomes incoherent when declared self-authorising — immune from the process that generated its authority — whether it is a sacred text, an institutional pronouncement, a moral code, or a scientific theory. Peirce's (1877) methods of tenacity and of authority are the third and fourth stages seen as methods of fixing belief, the one refusing the anomaly and the other institutionalising the refusal; the anatomy adds the two stages that precede them and the conflation that carries a standard from the first to the third.

What would refute the anatomy. The identity claim is what the anatomy stakes, and refutable. What would break it is a standard declared irrevisable that nonetheless retained its corrective process: an authority immune from challenge whose anomalies were still registered, whose dissenters were still heard, and whose foundation still answered to the evidence that produced it — Stage 3 without Stage 4, foreclosure without severance.

The anatomy predicts that this does not occur, a prediction that is no tautology, because the two stages are separately specified. Foreclosure is a claim about what may be challenged; severance is a claim about what the standard remains answerable to; nothing in the definition of the first supplies the second. A documented case of the one without the other would show the four stages to be one path among several rather than the structure of the pathology, so the anatomy would have to be rebuilt as a description rather than a mechanism.

One asymmetry among the domains belongs in the statement of the claim. In the scriptural and institutional cases Stage 2 is often declared;

in the scientific case it is enacted, the consensus treated as settled in the conduct of gatekeeping while no one announces its permanence. The identity across domains is functional, holding at the level of what the four stages do. Where the declaration is tacit the diagnostic is conduct: anomaly-dismissal, the marginalising of dissent, and severance from the corrective process identify Stage 2 by its downstream signature.

7 The Anatomy Applied

7.1 Slavery and the foreclosure of empathic recognition

The case of slavery exposes the four-stage mechanism with particular clarity and connects the anatomy directly to the great moral conflation of Chapter 16.

Stage 1. The moral code that permitted slavery in the antebellum American South was perfect in the technical sense. It was aligned with the prevailing standards — scriptural interpretation, economic analysis, legal precedent — and no challenge had successfully overturned it. The evidence available to the community converged on the correctness of the code.

“Perfect” here is the technical term of §4.1, complete alignment with the prevailing standards on the evidence in use, so the term must not be allowed to flatter the case. The standards were themselves corrupt, and evidence was available that the community refused to take up. The testimony of the enslaved was discounted, the credibility deficit Fricker (2007) names testimonial injustice, so the perfection was relative to an evidential base curated by the affected ignorance Stage 3 institutionalises (Moody-Adams, 1999). The technical sense marks how the code presented from within, unrefuted by everything admitted as evidence, without exculpating the community that did the admitting.

Stage 2. The perfection was elevated to infallibility: the code could not be wrong, because God had ordained it, because the scriptural basis was divinely inspired, and because the tradition had withstood all challenges.

Stage 3. New empathic evidence — the testimony of enslaved persons, the direct perception of suffering, the moral arguments of abolitionists — was structurally foreclosed. Not because the evidence was weak, but because the infallibility declaration placed the code beyond the reach of evidence.

The anchor result of Chapter 16 (§6.7) states what kind of thing was thereby barred. Not future arguments, but existing moral facts: the enslaved persons' own registrations, fully present and constituting the wrongness then and there (Chapter 14, §6.4). The declaration did not delay moral truth's arrival; it refused delivery of truth already at the door. Affected ignorance is the refusal of uptake; the anchor is what there was to take up. The abolitionists were not merely arguing against a strong consensus; they were arguing against a standard that had been declared unrevisable.

Stage 4. The moral code was severed from the empathic foundation that gave it authority. The code generated by empathic responses to suffering (the original prohibition against certain forms of cruelty) was no longer answerable to empathic responses to suffering (the recognition that enslaving persons causes suffering the original empathic responses would have condemned).

Stage 4 and Chapter 16's banking result (§6.7) are one mechanism at two settings of one switch. A revisable orthotic layer banks empathic corrections, which is why moral progress ratchets; a layer declared infallible embalms empathic errors, which is why moral regress locks in. The infallibility declaration is the ratchet run in reverse — the conversion of the instrument that holds corrections into the instrument that holds them out.

Nussbaum (2004) analyses the dehumanisation slavery required as the systematic suppression of empathic engagement, the deliberate refusal to register the enslaved person's suffering as morally relevant. On the present analysis the suppression was made possible by the infallibility declaration, which foreclosed the empathic evidence that would have overturned the code. Moody-Adams (1999) argues that moral progress is the overcoming of affected ignorance, the refusal to

acknowledge what one already has the resources to know, where the infallibility declaration is the institutional mechanism that sustains it.

The value taxonomy (Chapter 15) explains the resistance. The derived value of an infallible text — its association with God, sacred tradition, ancestral authority — overrode any primary-value response to the suffering the code permitted. The catastrophe was not merely that people held wrong beliefs. It was that the beliefs had been placed in a category that prevented the self-corrective mechanism from operating.

The same analysis applies elsewhere: to the subordination of women, the criminalisation of same-sex relationships, and every other case in which a moral code once considered permanently correct has been revised in light of deeper empathic engagement with the persons affected.

7.2 Three further domains

Institutional authority. When a religious or political authority declares its pronouncements beyond error, dissent becomes impossible within the system. If the institution cannot err, any disagreement is by definition the dissenter's error, so the capacity for self-correction is eliminated. Küng (1971) argued that papal infallibility is historically contingent, theologically unnecessary, and practically damaging; the present analysis provides the diagnosis his argument lacks, since the claim severs the orthotic standard from its empathic foundation, disabling the conditions for deep persuasion within the institution.

Moral codes. When a moral code is declared permanently correct, it is treated as infallible even if the word is not used, so the severance follows. Mill (1859/1999) identified the mechanism in his defence of free expression: even if an opinion is true, treating it as infallible prevents understanding of its grounds, reduces it to dead dogma, and eliminates the rational contestation through which truth is maintained.

Scientific consensus. The mechanism operates in science whenever the current consensus is treated not as the best available theory, which is perfection, but as permanently settled truth, which is infallibility. The decade-long dismissal of Marshall and Warren's (1984) identification of

Helicobacter pylori as the cause of peptic ulcers is the case usually cited, and §6's refutation condition is tried on it, since the consensus dismissed the anomaly for a decade and then corrected itself. No declaration of irrevisability stood behind the dismissal, and the ulcer consensus never severed itself from the corrective process, which is why the correction eventually ran. The case is therefore, on the anatomy's own terms, perfection correcting slowly — the ratchet turning late — and it earns its place here by marking the line: anomalous evidence explained away is Stage 3 conduct, and it becomes the pathology only where the corrective process that would eventually admit the anomaly is itself disabled. The anatomy's scientific instances are the cases where that disabling occurs; the ulcer case shows what the ratchet looks like when it does not.

Where the disabling has occurred, the case is Lysenkoism, and it walks the four stages in order (Joravsky, 1970). Lysenko's agronomy earned its standing on practical claims about the vernalisation of crops (Stage 1). The 1948 session of the Lenin Academy of Agricultural Sciences declared Michurinist biology the settled doctrine of Soviet science (Stage 2). Mendelian genetics was dismissed as bourgeois pseudoscience and its anomalies explained away (Stage 3). And the corrective process itself was dismantled — institutes closed, journals purged, geneticists dismissed and imprisoned, Vavilov among those who died in custody — so that revision became possible only when the political protection lapsed, in 1964 (Stage 4). The case confirms the anatomy in the one respect the ulcer case cannot, since the four stages appear in science as they appear in creed and code; and it confirms the qualification: reaching Stage 4 took an external authority to remove the correction science institutionalises, which is why the scientific instance is rare.

Oreskes (2019) analyses why scientific consensus deserves trust, since the social process of inquiry filters out errors over time, to which the distinction adds what her analysis needs. Justified trust in consensus is trust in perfection; elevation of consensus to infallibility forecloses the very social process that made the consensus trustworthy. The inclusion of scientific dogmatism is essential to the chapter's integrity,

because the mechanism operates identically across religious and secular institutions, targeting any standard, from any source, that claims immunity from the process that generated its authority.

7.3 The connection to the great moral conflation

The four-stage anatomy connects directly to the great moral conflation (Chapter 16). An infallible standard is an orthotic standard declared permanently fixed, and Stage 4 states the connection: severed from the empathic evaluations that give it authority, the infallible standard is a moral technology that has cut itself off from the source of its own legitimacy.

The anatomy also forecloses the conditions for deep persuasion (Chapter 21), derivably so. Deep persuasion is field-enrichment (Chapter 15, §4.4) applied to evaluative hierarchies, and an infallible standard is a field closed to enrichment by decree, one class of secondary object ruled inadmissible in advance. Deep persuasion requires that evaluative hierarchies be transformable through engagement. If the standard against which behaviour is judged is revisable, agents can debate not only whether behaviour conforms to the standard but whether the standard is adequate. If the standard is declared infallible, engagement is limited to the question of conformity, so the deeper question — is the standard itself good? — is foreclosed.

8 Objections and Replies

8.1 The moral stability objection

Objection. If no standard can be declared infallible, morality becomes a moving target.

Reply. The distinction does not undermine stability; it clarifies its nature. A perfect standard, one aligned with the best available evidence and the deepest available empathic insight, is the best we have, to be followed until a better alternative is demonstrated through evidence and persuasion. What the distinction undermines is false stability: the illusion that a standard is permanently settled when it is in fact

revisable. That is more dangerous than real revisability, because it prevents the correction of errors that may cause real suffering. The code permitting slavery was once considered as stable as any moral standard, and its revision was one of humanity's great moral achievements.

8.2 The self-undermining objection

Objection. If no claim is infallible, then the claim that no claim is infallible is itself revisable.

Reply. The chapter's central claim, that perfection does not entail infallibility, is a structural result rather than an empirical generalisation. The inference from a finite body of evidence to a universal claim about all possible evidence is invalid because the space of possible evidence is open-ended. This is a claim about the logical relationship between finite and infinite domains, not an inductive claim that might be overturned by a counterexample.

The distinction between the two routes provides the basis. Route 1 is the omniscient guarantee, structurally unavailable; Route 2 is logical self-vindication, available for self-defeating denials. The claim that perfection does not entail infallibility is not self-vindicating, since its denial does not instantiate what it denies, yet it is established by structural argument rather than by inductive generalisation. It is perfect, not infallible — in the Register's terms, established (§4.3), a tier defined as a species of perfection. The framework's consistency here is therefore a correspondence built into the grading system itself rather than a happy accident to be checked case by case.

8.3 The religious authority objection

Objection. For many believers, the authority of scripture is a direct conviction grounded in faith rather than an inference from perfection.

Reply. The chapter does not deny that the conviction is sincerely held or grounded in a powerful experience of the sacred. A person can maintain a profound, reverent, and deeply engaged relationship with a text while recognising it as excellent — possibly the best moral technology available — without declaring it incapable of error. The

classification diagnoses the conviction: the felt impossibility of error is on the emotional-pull axis, sustained by the derived value of the text's sacred associations, not on the logical axis. The conviction is real; it is not evidence that the text is infallible.

8.4 The practical indispensability objection

Objection. Communities need stable foundations. Declaring core standards revisable invites endless contestation.

Reply. The practical benefits of stability are real, and perfection provides them without infallibility. A community that treats its best standards as excellent and follows them with conviction has all the stability infallibility promises. What infallibility adds is not stability but unreformability: the foreclosure of the process through which the standards could be improved. The history of moral progress is a history of communities revising standards they once treated as beyond challenge, so practical stability requires confidence in one's standards rather than a declaration that they are incapable of error.

8.5 The bypass objection

Objection. The chapter's argument targets the inference from perfection to infallibility. But infallibility need not be inferred: it can be established by testimony (the prophet reports God's word), by religious experience (the believer feels the text's divine origin), or by properly basic belief (the sensus divinitatis warrants belief without inference).

Reply. Each bypass fails for the same reason: none can establish the universal claim infallibility requires. Testimony transmits knowledge under appropriate conditions (Lackey, 2008), but a finite testifier cannot confirm an infinite claim. Religious experience is evidence that an experience occurred (a real and potentially transformative evaluative event), not that its cause is omniscient. Properly basic belief is addressed at §5.2: warrant is defeasible on Plantinga's own account, so proper basicality delivers perfection and not infallibility. Each strategy may produce perfection — justified, well-grounded, even excellent belief. None produces infallibility.

A fourth bypass is the classical one, met elsewhere: omniscience is not verified at all but entailed, either by the concept of a perfect being or by the cosmological route to a first cause. Section 5.1 does not touch this, and does not need to. The entailment carries only if its terminus can bear cognition at all, which Chapter 10 (§3.2) denies at theorem strength, and only if the cause can be shown necessary, which Chapter 11 denies it can. The route is closed upstream rather than here.

9 Conclusion

The chapter yields three results.

The two-dimensional classification (established). Crossing a logical axis with an emotional-pull axis generates four epistemic states that explain dogmatic closure, sceptical paralysis, together with the phenomenological indistinguishability of grounded and ungrounded conviction. Irrational certainty is the epistemological basis of dogma.

Perfection is not infallibility (established). Perfection is an evidence-based, standard-relative, temporal achievement; infallibility is a universal claim about all possible evidence, which no finite body of evidence can establish. The impossible-evidence argument shows that only omniscience can verify omniscience. The two-infinities argument shows that infallibility's metaphysical precondition, omniscience conjoined with divine simplicity, is structurally incoherent. Infallibility is therefore unsupported for any guarantor whatever, by the evidential argument, and unavailable for the guarantor classical theism conceives, by the structural one — a theism that relinquishes divine simplicity escaping the second and not the first.

The anatomy of dogma (constructed). The four stages — earned authority, extended permanence, foreclosed revision, severed foundation — are the mechanism by which revisable standards become unreformable, operating identically across scripture, institutional authority, moral codes, and scientific consensus. An infallible standard is an orthotic standard severed from the empathic foundation that gives it authority, so the severance forecloses the conditions of deep persuasion (§7.3).

If moral standards are revisable rather than infallible, the agent's relationship to those standards is one of engagement rather than submission. But the agent's capacity for engagement is itself limited. The evaluative hierarchy (Chapter 15) from which she evaluates standards was not itself chosen; it was delivered by processes she did not command. Chapter 18 argues that intentional and unintentional behaviour share the same chain architecture — a unified causal model in which the evaluative hierarchy that determines choice was not itself chosen — and that this holds for any possible mind, in any possible world.

Grades and Dependencies

1. **The anatomy's stake.** The classification and the perfection/infallibility distinction are *established*; the four-stage anatomy is *constructed*, its identity claim refutable (§6): documented Stage 3 without Stage 4 demotes it from mechanism to description; tacit Stage 2 is diagnosed by conduct. The two-infinities argument is scoped (§5.2): structural against the classically simple guarantor, evidential against every guarantor; the grain condition (Ch. 2) attaches to the content-level formulation only. The Premise Register carries the chain.

References

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, 275(5304), 1293–1295.

Brady, M. S. (2013). *Emotional insight: The epistemic role of emotional experience*. Oxford University Press.

Cassam, Q. (2019). *Vices of the mind: From the intellectual to the political*. Oxford University Press.

Christensen, D. (2007). Epistemology of disagreement: The good news. *Philosophical Review*, 116(2), 187–217.

Coliva, A. (2015). *Extended rationality: A hinge epistemology*. Palgrave Macmillan.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. Putnam.

De Sousa, R. (1987). *The rationality of emotion*. MIT Press.

Elga, A. (2007). Reflection and disagreement. *Noûs*, 41(3), 478–502.
Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, 3(4), 552–564.
Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.
Friedman, J. (2019). Inquiry and belief. *Noûs*, 53(2), 296–315.
Harman, G. (1973). *Thought*. Princeton University Press.
Hookway, C. (2008). Questions, epistemology, and inquiries. *Grazer Philosophische Studien*, 77(1), 1–21.
Hume, D. (1748/1999). *An enquiry concerning human understanding* (T. L. Beauchamp, Ed.). Oxford University Press.
Joravsky, D. (1970). *The Lysenko affair*. Harvard University Press.
Kelly, T. (2010). Peer disagreement and higher-order evidence. In R. Feldman & T. Warfield (Eds.), *Disagreement* (pp. 111–174). Oxford University Press.
Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, 100(4), 609–639.
Kripke, S. A. (2011). On two paradoxes of knowledge. In *Philosophical troubles: Collected papers* (Vol. 1, pp. 27–51). Oxford University Press.
Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.
Küng, H. (1971). *Infallible? An inquiry* (E. Quinn, Trans.). Doubleday.
Lackey, J. (2008). *Learning from words: Testimony as a source of knowledge*. Oxford University Press.
Locke, J. (1689/1975). *An essay concerning human understanding* (P. H. Nidditch, Ed.). Clarendon Press.
Maia, T. V., & McClelland, J. L. (2004). A reexamination of the evidence for the somatic marker hypothesis: What participants really know in the Iowa gambling task. *Proceedings of the National Academy of Sciences*, 101(45), 16075–16080.
Marshall, B. J., & Warren, J. R. (1984). Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet*, 323(8390), 1311–1315.
Mill, J. S. (1859/1999). *On liberty* (E. Alexander, Ed.). Broadview Press.
Moody-Adams, M. M. (1999). The idea of moral progress. *Metaphilosophy*, 30(3), 168–185.
Nussbaum, M. C. (2004). *Hiding from humanity: Disgust, shame, and the law*. Princeton University Press.
Oreskes, N. (2019). *Why trust science?* Princeton University Press.
Peirce, C. S. (1877). The fixation of belief. *Popular Science Monthly*, 12, 1–15.
Plantinga, A. (1980). *Does God have a nature?* Marquette University Press.
Plantinga, A. (2000). *Warranted Christian belief*. Oxford University Press.

Pollock, J. L. (1986). *Contemporary theories of knowledge*. Rowman & Littlefield.

Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. Routledge.

Pritchard, D. (2016). *Epistemic angst: Radical skepticism and the groundlessness of our believing*. Princeton University Press.

Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review*, 60(1), 20–43.

Smithies, D. (2019). *The epistemic role of consciousness*. Oxford University Press.

Sosa, E. (2007). *A virtue epistemology: Apt belief and reflective knowledge* (Vol. 1). Clarendon Press.

Stanley, J. (2005). *Knowledge and practical interests*. Clarendon Press.

Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.

Wittgenstein, L. (1969). *On certainty* (G. E. M. Anscombe & G. H. von Wright, Eds.; D. Paul & G. E. M. Anscombe, Trans.). Blackwell.

CHAPTER 18

Free Will and the Perspectival Asymmetry

Presupposes: the categorical transition principle (Ch. 13); the response-constitutive account of value (Ch. 14); the primary/added value taxonomy (Ch. 15).

Establishes: the unified causal model, for any possible substrate; the perspectival asymmetry, with the two sources of felt openness it separates; bounded authorship and its consequence for moral responsibility.[2]

In brief. A rock rolls downhill because gravity pulls it. You chose your career because something pulled you too. The forces differ — one gravitational, one emotional — but the structure is the same: a chain of attractions and repulsions producing an outcome the final link did not author. The free will problem, this chapter argues, is not really about determinism but about self-consciousness. In the moment of choosing, the chooser and the choice are identical — so freedom is undeniable from inside; from the retrospective view, the chain that delivered your evaluative field is equally undeniable. Both reports are accurate; neither is complete. The result is *bounded authorship*: a musician improvising in a key she did not choose is a real author of her melody. Practical responsibility survives intact. What falls is the theological grade — desert sufficient for eternal sanction — and it falls for any possible mind, in any possible world.

---

1 Introduction

A rock rolls downhill because gravity pulls it. You chose your career because something pulled you too. The forces differ, one gravitational and the other emotional, while the structure is the same: a chain of attractive and repulsive forces producing an outcome neither the rock nor you authored from scratch. This chapter argues that the structural identity between these two cases is a philosophical finding with consequences for free will, moral responsibility, and the theology of divine punishment.

The free will debate has been framed, for centuries, as a question about determinism.

Compatibilists (Frankfurt, 1971; Dennett, 1984; Fischer & Ravizza, 1998) argue that freedom is compatible with determinism, provided the agent's actions flow from the right kind of internal states. Libertarians (Kane, 1996; O'Connor, 2000) argue that real freedom requires metaphysical indeterminacy. Hard incompatibilists (Pereboom, 2001, 2014) argue that free will is an illusion regardless of whether determinism is true. The debate turns on whether the causal structure of the world leaves room for agency, and, if it does, what kind of agency survives. The reactive-attitudes tradition (P. F. Strawson, 1962) reframed those terms by relocating responsibility in the attitudes of interpersonal life rather than in a metaphysical condition; the compatibilist accounts above inherit that relocation, while the chapter's verdict keeps the practice apart from the desert reading of it.

This chapter does not enter the debate on those terms. It argues that the free will problem is not fundamentally about determinism — about whether the causal chain is open or closed — but about the structural relationship between consciousness and the content it operates on. The chapter develops three findings that the existing literature does not combine and that, taken together, constitute a contribution the established frameworks do not jointly provide.

The first finding is the unified causal model. All behaviour, intentional and unintentional, results from a chain of attractive and repulsive forces: physical for rocks, stars, and machines; physical plus emotional for sentient beings, whose evaluative modality constitutes oriented attractions and repulsions (Chapter 14). The addition of emotion changes what the chain can produce — evaluation, choice, purpose — without changing its fundamental structure: a sequence of causes delivering an outcome the final link did not author. The model's key term is defined rather than borrowed, since §2.2 gives an evaluative force a direction, a magnitude, and a composition rule, and states where the physical comparison holds and where it ends.

The second finding is the perspectival asymmetry. "Freedom is real" and "freedom is illusory" are complementary reports from two structurally asymmetric perspectives rather than competing descriptions of one phenomenon. From the first-person present, freedom is

structurally undeniable, because the chooser and the choice are identical: disagreeing with your current orientation would require a different orientation, which would then be yours. From the third-person retrospective perspective, the determination is equally undeniable, because the choice was the product of a chain the agent did not initiate.

This is not a restatement of compatibilism. The asymmetry holds that the first-person perspective delivers an accurate report of its own condition — the agent really is evaluating — while structurally lacking access to what the third-person perspective reveals, that the evaluation was determined. Neither perspective is wrong; each is incomplete.

The third finding is the ethereal mind argument: both results hold for any possible conscious system — physical, non-physical, computational, or otherwise — because they follow from the structure of evaluation and self-consciousness rather than from the composition of brains. Even granting substance dualism, libertarian agent causation, and metaphysical indeterminacy, the first content to populate any conscious field cannot itself be chosen, and the evaluative responses that constitute preference are delivered by processes outside the control centre.

The three findings jointly produce what this chapter calls bounded authorship: the agent a real author at the phenomenological level, where her choices flow from a hierarchy that is hers, and no author at the metaphysical level, where the hierarchy was delivered by a chain she did not initiate. §6.1 states the concept in full.

A note on the existing literature. That subconscious processes shape our choices is well-documented empirically (Kahneman, 2011; Damasio, 1994) and developed philosophically by Frankfurt (1971, 1988), Dennett (1984, 2003), and Mele (2006); practical responsibility without ultimate origination is a position occupied by Mele, Vargas (2013), and Fischer and Ravizza (1998).

The chapter claims no priority on the shared ground. What it claims is that the three findings above are contributions the existing literature does not provide, and that their conjunction is not replicated by any single framework, which §6 develops through the convergences and

divergences. Stated at its narrowest, the dispute with compatibilism concerns desert (§8.1), and the three findings supply the mechanism and the substrate-generality behind that verdict.

A clarification developed at §7.2 should govern the reading from the start. The "absolute" responsibility the chapter denies is the theological grade — desert sufficient to justify unbounded or eternal sanction — while the practical responsibility on which human life and law run is preserved throughout.

The chapter develops the unified causal model (§2), the delivery of evaluative content through the subconscious jump and the bounded control centre (§3), the perspectival asymmetry (§4), the ethereal mind argument (§5), the relationship to four established positions in the free will debate (§6), and the consequences for moral responsibility (§7).

2 The Unified Chain

2.1 Attraction, repulsion, and the common structure

Consider a rock dislodged by wind. The wind exerts a force; the rock moves, strikes another rock, falls into a river. At every stage the outcome is determined by the balance of attractive and repulsive forces — gravity, friction, current — and the chain is entirely unintentional, since nothing within it evaluates, prefers, or chooses.

Now consider a person deciding between two careers, one financially secure but unfulfilling, the other risky but engaging. She deliberates and chooses the risky one. At every stage the outcome is determined by the balance of attractive and repulsive forces — the pull of engagement, the push of financial anxiety, the pull of self-expression, the push of social pressure — and the chain is intentional, since the person evaluates, prefers, and chooses. The outcome is the product of forces acting on a system with oriented responses to them.

The unified causal model identifies the structural identity between these two cases, and once §2.2 has disclaimed the physics, the identity is that both are sequences of causes producing an outcome the final link did not author, which describes the cases. Both are chains of attractive

and repulsive forces producing an outcome. The difference is not in the chain's structure but in the kind of force. Physical behaviour is governed by forces that produce motion without evaluation: gravitational, electromagnetic, nuclear. Intentional behaviour is governed by the same physical forces plus the evaluative modality's oriented responses: the emotional attractions and repulsions that Chapter 14 identified as constitutive of evaluative difference. The evaluative modality adds a new kind of force to the chain; it does not change the chain's architecture.

Kim (1998) develops the most rigorous account of the relationship between mental and physical causation, arguing that mental properties are causally efficacious only if they are identical with or realised by physical properties: the problem of causal exclusion. The unified model engages Kim at a specific point, since the evaluative modality's oriented responses are forces within the causal chain rather than epiphenomenal accompaniments to it.

Whether these forces are identical with physical processes, as the physicalist holds, realised by them, as the non-reductive physicalist holds, or constituted by a non-physical substance, as the dualist holds, their role in the chain is the same: they are among the attractive and repulsive forces that determine the outcome. The unified model is therefore neutral on the metaphysics of mind-body interaction, identifying the role the evaluative forces play regardless of their ontological status.

Anscombe (1957) argued that intentional action is distinguished from mere bodily movement by its susceptibility to a specific form of explanation: explanation in terms of the agent's reasons. Davidson (1963) developed this into the thesis that reasons are causes, so that the agent's beliefs and desires causally produce her actions, so the rationalising explanation of an action just is its causal explanation. The unified model agrees with Davidson that reasons are causes while identifying the specific mechanism: the "reasons" that cause intentional action are the evaluative modality's oriented responses, emotional attractions and repulsions that enter the causal chain through the

subconscious jump (§3). What makes a cause a “reason” is the presence of the evaluative modality in the chain.

2.2 The comparative model

The unified model can be stated as a comparative thesis. In unintentional behaviour, the chain runs: external stimulus → physical forces (attraction/repulsion) → outcome. In intentional behaviour, the chain runs: external or internal stimulus → physical forces + evaluative forces (emotional attraction/repulsion) → outcome. The two chains share three features.

First, both are chains: sequences of causally connected events in which each link is produced by the preceding one. The rock is caused to fall by gravity and wind; the person’s preferences are caused by the evaluative modality’s response to her developmental history, circumstances, and subconscious associations.

Second, both are governed by the balance of attractive and repulsive forces: the rock’s trajectory by the resultant of gravity, friction, and wind; the person’s choice by the resultant of competing emotional responses, the composite of primary and added value of Chapter 15. Third, in neither case does the final link author the chain, since the rock did not create the wind and the person did not create the responses that constitute her preferences. Both are endpoints of chains they did not initiate.

The difference between the two cases is that the intentional chain includes the evaluative modality, the system that constitutes oriented differences between better and worse, desirable and undesirable (Chapter 14). This addition is what makes intentional behaviour intentional, since it introduces evaluation, comparison, preference, and choice.

But the addition does not change the chain’s structure. It adds a new kind of force, emotional rather than physical, without altering the architecture of causation. The person’s choice is as much a product of the forces acting on her as the rock’s trajectory is a product of the forces

acting on it. What differs is the kind of forces doing the determining rather than the presence or absence of determination.

What "force" means here. The word "force" invites a challenge that should be met where the word is introduced. Either it is a metaphor borrowed from physics, in which case the model decorates a familiar claim, or it means only that mental events have causes, in which case the model asserts what every naturalist already grants.

The reply is that the term is defined, so the definition gives the model its content. An evaluative force is an oriented response with a direction and a magnitude. The direction is its valence, the toward or away, the directional difference of Chapter 13 (§2.3). The magnitude is the intensity of the response, the quantity the value taxonomy operates on, and the one the forced-choice diagnostic makes evidentially accessible (Chapter 15, §5.2). Chapter 15 supplies the composition rule: the total value of an option is the composite of its primary and added components, and choice follows the strongest total value.

The comparison with physics is accordingly a structural analogy, holding at three points and silent beyond them. The components have direction and magnitude; they compose into a resultant; the resultant fixes the outcome. Beyond those three points the analogy ends, since evaluative forces occupy no physical space, obey no conservation law, and share no unit with mechanics. The model claims the shared form and nothing further, a form that earns its keep downstream: the resultant/component distinction carries the reply to Frankfurt's unwilling addict, and the composition rule is what the perspectival asymmetry describes when it speaks of the field's total upshot (§4.1).

Schopenhauer (1839/1985) anticipated the unified model's central insight: a man can do what he wills, but he cannot will what he wills. The observation identifies the limit without providing its mechanism. The unified model provides the mechanism, since the will is constituted by the evaluative modality's oriented responses, which are links in a causal chain the agent did not author. The agent can act on her strongest response, so she can do what she wills; she cannot choose which response is strongest, so she cannot will what she wills, because the

responses are delivered by the chain rather than produced by the control centre.

3 The Subconscious Architecture

3.1 The subconscious jump

Where do the evaluative forces that populate the intentional chain come from? From processes consciousness does not control.

Consider a simple case. You are trying to remember someone's name. You focus on their face, on a conversation you had, on the context in which you met. You are casting bait into the subconscious, stimulating associative networks in the hope that the right item will surface. The name may appear suddenly and unbidden, or it may not appear at all. Either way, the delivery of the name to consciousness is not a conscious act. You did not choose to remember it at that moment; it arrived.

The specific moment at which content transitions from subconscious processing to conscious awareness is the subconscious jump. Three features of the jump matter for the unified model.

First, it is involuntary: we do not choose which thoughts surface, which emotions emerge, or which memories are recalled. We can cast a line into the ocean of the subconscious, imagining which fish might bite, without deciding which fish will jump out.

Second, it is constitutive of the evaluative field. The emotions, memories, and thoughts that arrive constitute the raw material of evaluation, and we do not first choose our emotional responses and then evaluate; they arrive, and we evaluate with and through them. Third, it is structurally prior to choice, since the emotional responses that constitute a preference (Chapter 15) are delivered before the choice is made. By the time conscious deliberation begins the field has already been populated, so deliberation operates on this content rather than producing it.

3.2 The bounded control centre

Consciousness can be understood as the control centre of the mind: the domain in which deliberate evaluation occurs. Within this domain, a subject can attend to specific thoughts, weigh options, compare alternatives, and arrive at a choice. But the control centre is structurally bounded. Everything outside it (the body, the subconscious, hormonal systems, neural micro-processes) influences the evaluative field without being directed by deliberate evaluation. The relationship is asymmetric: subconscious processes shape conscious content, but consciousness does not shape subconscious processes in the same direct way.

The boundedness can be illustrated through a series of progressively deeper influences on the evaluative field. External coercion (a knife at one's back) constrains the options available but is recognisable as external. Drugs, psychiatric conditions, or brain tumours alter the evaluative field from within, affecting personality and decision-making through chemical or physical changes the agent may not detect. Hormonal shifts produce emotional states (anger, anxiety, desire) that the agent experiences as her own orientations without having chosen them. And at the deepest level, the micro-constituents of consciousness (the neural or subatomic processes that constitute thought) are entirely outside the control centre's jurisdiction. A change at this level alters the evaluative field without the agent's awareness that any change has occurred.

The control centre is therefore bounded on all sides: from above by external circumstances, from below by physiological processes, and from within by the subconscious delivery of evaluative content. Deliberation is real — the agent does weigh options, compare outcomes, and select a course of action — while operating on material it did not produce, within a field it did not populate, using responses it did not author.

3.3 The subconscious hierarchy

The evaluative hierarchy has a subconscious counterpart, set out in Chapter 15 (§5.3). Emotional patterns stored beneath awareness —

associations with memories, fears, desires, and traumas — shape thoughts, preferences, and actions in ways the conscious mind neither recognises nor authored, forming the involuntarily constituted base on which the conscious hierarchy is built.

A person who inexplicably avoids a particular kind of social situation may be responding to a subconscious association between that situation and a childhood experience of humiliation: a derived-value connection (Chapter 15) stored beneath awareness, powerful enough to shape the conscious hierarchy. Psychological analysis aims to uncover these hidden layers — forgotten or unformed associations that guide behaviour without announcing themselves.

The unified model's point can now be put sharply. The forces that determine intentional behaviour are delivered by the subconscious jump, shaped by a subconscious hierarchy the agent did not author, and processed within a control centre bounded on all sides. The agent deliberates in earnest, with material she did not choose, within a chain she did not initiate.

4 The Perspectival Asymmetry

4.1 Why freedom feels real

If the evaluative field is involuntarily constituted and the chain of causes extends back beyond the agent's control, why does freedom feel so vivid and immediate? The question has received three principal answers in the literature. Dennett (1984, 2003) argues that freedom feels real because it is real: the agent possesses the kind of freedom worth wanting, and the feeling accurately tracks this. Pereboom (2001, 2014) argues that freedom feels real despite being illusory, so the phenomenology is misleading. Wegner (2002) argues that the experience of willing is a post-hoc confabulation, the feeling of authorship generated after the neural processes that produced the action have already completed.

The perspectival asymmetry offers a fourth answer, distinct from all three. Freedom feels real because of a feature of self-consciousness that

none of the existing accounts identifies: in the present moment of choosing, the chooser and the choice are identical.

The argument can be stated in one move. When you are choosing — when the evaluative process is actively weighing options and orienting toward one over another — you cannot simultaneously disagree with the orientation. To disagree with your current preference would require a different preference to disagree from, and that different preference would then be your current one.

You can disagree with a past choice, viewed retrospectively from the third-person perspective. You can anticipate disagreeing with a future choice, viewed prospectively. But you cannot disagree with the choice you are making right now, because the choice you are making right now just is the current state of your evaluative field, and you are that field.

This is not a psychological observation about the limits of introspection. It is a claim about the relationship between the evaluative field and the self that operates within it — a picture of the self, inherited from Chapter 13 (Level 4), that a stronger conception of the reflective agent denies; the denial is met below. The self-conscious system (Chapter 13, Level 4) is not a homunculus sitting behind the evaluative field and inspecting it from outside. It is the evaluative field's own self-awareness, the field detecting itself as a unified, oriented entity.

The "I" that chooses is therefore not separate from the hierarchy that constitutes the choice; the "I" is the hierarchy, experienced from the inside. And a hierarchy cannot disagree with its own current state, because disagreement is itself an evaluative orientation that would become the new current state. By itself this establishes only that present-tense disagreement is unavailable. That the determination is also inaccessible from within — that the agent cannot see the chain that produced her current orientation — is the further claim argued below (§4.2); it does not follow from the identity alone.

The unwilling addict. An apparent counterexample sharpens the claim rather than defeating it. Frankfurt's (1971) unwilling addict acts on a desire he simultaneously repudiates: the craving carries him to the

drug while he condemns the craving. This looks like present-tense dissent from his own effective motivation, and akrasia generally appears to exhibit the same structure.

The reply turns on what "the choice" names. The choice, on the present account, is the resultant of the evaluative field — the outcome of the totality of oriented responses, as the unified model of §2.2 has it — rather than any component within the field. The addict's repudiation is real; it is inside the field: a negative evaluative response taking his own craving as its object, one vector among the vectors whose resultant carries him to the drug.

What the perspectival identity rules out is not dissent from a component, which is common and is what conflict within the hierarchy consists in (Chapter 15, §4), but dissent from the resultant as resultant. A further orientation against the field's total upshot would be a further vector, already part of the total it purports to oppose, altering the resultant if strong enough and absorbed into it if not. The addict confirms the analysis: his condemnation, had it been stronger, would have been the choice, and being weaker it is a component of the choice he laments. Retrospectively he can disagree with what the field delivered, the third-person perspective doing what §4.2 describes, while in the moment the disagreeing and the being-carried are both his current field; there is no position within the field from which its own resultant can be opposed.

Korsgaard (1996b, 2009) develops the most sustained philosophical account of the first-person standpoint as constitutive of agency. She argues that the deliberative perspective, the perspective from which we ask "what shall I do?", is irreducible to the theoretical perspective from which we observe causal regularities. On Korsgaard's account, the agent who deliberates is not applying a theory of action to herself; she is exercising a form of self-constitution available only from the first-person standpoint.

The asymmetry converges with Korsgaard on the irreducibility of the first-person standpoint and diverges on its implications. Korsgaard treats the standpoint as constitutive of real freedom, so the deliberating

agent is self-governing. The asymmetry holds that the standpoint delivers an accurate report of its own phenomenological condition while unable to access what the third-person perspective reveals. Korsgaard's self-constitution is therefore real at the phenomenological level and bounded at the metaphysical one, the two claims describing the same event from perspectives self-consciousness makes necessarily divergent.

Korsgaard (2009) supplies the position with a regress-stopper the engagement must meet, because it is aimed at arguments of this chapter's shape. On the constitutivist account, the principles of practical reason stand apart from the items an agent might endorse or decline: they are constitutive of agency itself, so asking for their credentials is asking from nowhere. An agent who deliberates at all is already committed to them, and the regress of endorsements (Watson's question, §6.2) terminates where agency begins.

Three replies locate the present view. The first concerns delivery. Whatever the constitutive principles' status, their grip on a particular agent — the felt claim of unity, the pull toward being one practical identity — arrives through the same subconscious channel as every other orientation (§3.1), so the constitutivist has described the content of the endowment while leaving its delivery where the chapter places it. A principle constitutive of agency is still a principle the agent received with her agency.

The second is the resultant point applied to the deliberative standpoint itself. The standpoint from which Korsgaard's agent constitutes herself is the evaluative field in operation, and no position within it permits dissent from its own resultant, so the self-constitution is real at the level the asymmetry assigns to it and bounded at the level the retrospective view discloses. Enoch (2006) presses the constitutivist from a third flank, since an agent may ask why she should care about what is constitutive of agency, and the aim's constitutive status supplies no answer. The present argument is compatible with either outcome of that exchange, because its claim concerns where the aim's grip comes from, whichever normative standing the aim is assigned.

A third reply is owed to the denial of the picture itself. On the field model the "I" is the hierarchy experienced from within, and present-tense dissent from the resultant is impossible because the dissenting is one more vector. A Korsgaardian or Frankfurtian holds that the reflective agent stands to her desires as endorser to endorsed, and that the standpoint of endorsement is the field's governor; on that picture present-tense dissent is what identification is. The account does not assume the denial away. It asks what the endorsing consists in, and gives the endorser two options: either the endorsement is an occurrence with a felt direction — a preference for this desire over that — in which case it is in the field, on the account's side of the picture, however governing it feels; or it is a registration that a relation obtains without any direction, in which case it moves nothing and cannot be the identification the picture needs (Chapter 14, §3.5). The picture of a governor outside the field survives only if there is a third thing for the governing to be, which is the question the grip argument of Chapter 16 (§3.1) puts to every requirement — felt, or mere content — answered here as it is answered there. That is the reply; it converts the denial into a production challenge, and meeting the challenge refutes the model; the argument built on it then stands or falls on its own.

4.2 The two perspectives

The perspectival asymmetry is a structural consequence of self-consciousness. Any self-conscious system — any system that has crossed the categorical transition to Level 4 (Chapter 13) — possesses two perspectives on its own choices.

The first-person present perspective. The system is currently evaluating, its evaluative modality constituting oriented differences between options, with the strongest orientation determining the choice. From this perspective the choice is experienced as the system's own, because the orientation is constituted by the system's own evaluative responses. The system experiences itself as the source of the choice, since the choice just is the current state of the evaluative field, and the system just is the field aware of itself.

The third-person retrospective perspective. The system reflects on a choice already made, or observes another system's choice. From this perspective the chain is visible: the choice was produced by evaluative responses delivered by the subconscious jump, shaped by developmental history, constituted by the evaluative modality's involuntary reactions, and situated within a hierarchy the agent did not author. The determination is clear, because the retrospective perspective has access to the causal history that the present-moment perspective lacks.

The asymmetry between these two perspectives is not a deficiency in either. It is a consequence of what it is to be a self-conscious evaluative system. The first-person present perspective cannot access the causal history of its own current state, because accessing that history would require stepping outside the evaluative field, which is the one thing this perspective cannot do, since it is the field. The third-person retrospective perspective cannot access the phenomenological reality of present-moment evaluation, because that reality is available only to the system currently doing the evaluating.

Two supports keep the inaccessibility claim from resting on assertion. The first is structural: the delivery of the field's contents is not an event in the field. What arrives in consciousness is the evaluative response — the pull, the aversion, the salience — never the constituting of the response, because a constituting that appeared would be one more delivered content, not the delivery witnessed (§3.1; Chapter 15, §5.3). Introspection can inspect the field's contents to any depth; it cannot inspect the production of the field, for the same reason a stage can present any scene except the machinery at the moment of raising it into view.

The empirical corroboration. The classic dissociation studies found subjects fluently reporting plausible causes of their preferences and choices while remaining unaware of the manipulated factors that demonstrably produced them: confabulation just where access fails (Nisbett & Wilson, 1977).

The sharper instance is choice blindness. Subjects who select one of two faces as more attractive, are covertly handed the rejected one, and are asked why they chose it, mostly fail to notice the substitution and supply reasons for a choice they did not make (Johansson, Hall, Sikström, & Olsson, 2005). Their reports for the manipulated trials are, by semantic analysis, barely distinguishable from their reports for the unmanipulated ones (Johansson, Hall, Sikström, Tärning, & Lind, 2006). That is the prediction at its most exposed: full access to the field's contents, none to their production, and no internal mark separating the two.

The effect is context-dependent, since cues that the experimenter is unreliable raise detection, which locates it as a fact about spontaneous monitoring rather than about the limits of scrutiny — and that is the fact the argument needs. It is the signature the structural argument predicts: a system with full access to its field's contents and none to their production, filling the gap with the nearest third-person theory to hand.

A distinction is being relied on here that should be named: between a content and its manifestation. The thesis of determination can itself enter the field, as a belief, one more delivered content, without the delivery thereby becoming witnessed. What the field can contain is the proposition that production occurs, never the production occurring. The line is not local to choice: what a system does is never among the things it registers, at any level of the hierarchy (Chapter 13, §2.5), and the perspectival asymmetry is that general asymmetry met at the one level where the doing is a choosing. The corroboration is not proof, and the scope of introspective access remains empirically contested, while the direction of the evidence is the direction the argument requires.

The asymmetry explains the persistence of the pre-theoretical debate. Each side of the folk dispute describes one perspective and correctly reports what is real within it — freedom from the first-person present, determination from the third-person retrospective — then mistakes the perspective for the whole truth. Within the professional literature the disagreement survives the asymmetry in a narrower form and concerns desert; §8.1 records that sophisticated compatibilists hold both descriptions and locate responsibility inside the second.

Nagel's (1986) distinction between the subjective and objective standpoints is the closest existing analogue to the perspectival asymmetry. Nagel argues that the subjective standpoint, the view from within a particular perspective, and the objective standpoint, the view from nowhere, are both real and irreducible, and that the tension between them is a permanent feature of the human condition.

The perspectival asymmetry agrees with Nagel's irreducibility thesis while providing a more specific account. The two perspectives are not merely different viewpoints on the same reality but perspectives whose divergence is explained by a specific feature: the identity of the chooser with the evaluative field in the present moment, which makes the first-person perspective incapable of accessing what the third-person perspective reveals. Nagel identifies the tension; the perspectival asymmetry identifies its mechanism.

4.3 The consequence for the free will debate

The perspectival asymmetry does not so much dissolve the free will debate as explain its intractability, and locate where its single metaphysical answer lies. The question that divides the tradition — is the agent the ultimate author of her choices? — has one answer, the third-person one: she is not. Sized accordingly, the asymmetry's contribution is the application to choice of a general fact about registration — that what a system does is never among the things it registers (Chapter 13, §2.5) — with the identity of chooser and choice as the reason the application bites here in a way it does not elsewhere; the position the mechanism reaches is Schopenhauer's; what it adds is the mechanism.

What the asymmetry adds is an account of why that answer is so hard to credit from the inside. "Freedom is real" and "freedom is illusory" are reports from two vantage points rather than competing reports of the same fact. The first-person present perspective accurately registers that the agent is evaluating, while structurally lacking access to the determination that the third-person retrospective perspective records. The experience is veridical as far as it reaches, since the agent really is evaluating. It does not reach the authorship. That is why the debate has

resisted resolution, each side having mistaken its vantage point for the whole.

The practical consequence is that the level of description that matters depends on the question being asked. For everyday accountability — for legal systems, for interpersonal relationships, for self-governance — the first-person present perspective is the operative one. The agent's choices reflect her evaluative hierarchy; she is responsible for them in the practical sense that the choices express who she is.

For ultimate accountability — for the theology of divine retribution, where eternal consequences are imposed on the basis of the agent's ultimate authorship of her choices — the third-person retrospective perspective is the operative one. The agent did not author the hierarchy, so holding her ultimately responsible for its outputs is holding her responsible for the product of a chain she did not initiate.

This two-level analysis — practical responsibility at the phenomenological level, metaphysical determination at the retrospective level — is what makes the theological application in §7 possible, and what distinguishes bounded authorship from the established positions (§6). Both results have been developed without specifying the substrate of consciousness. The standard debate is conducted within broadly physicalist assumptions, with libertarians invoking non-physical substance or agent-causal power to create space for indeterminacy. The next section shows that the structural constraint holds regardless of which metaphysics of mind is correct.

Where the argument stands: the unified causal model has been stated and its key term defined (§2); the delivery of evaluative content through the subconscious jump has been described (§3); the perspectival asymmetry has explained why determination is invisible from within (§4). What remains is to show that none of this depends on what minds are made of.

5 The Ethereal Mind Argument

5.1 The argument from initial content

The argument can be stated in a single sentence: we cannot choose what populates the origin of our mind. Choice is a second-order operation, since it operates on content that must already be present. Before the first thought, the first emotion, the first detected distinction, there is nothing to choose with and nothing to choose between. The initial populating of the conscious field is therefore involuntary; this holds regardless of what the mind is made of.

To see why, grant the libertarian everything she wants. Suppose substance dualism is true: the mind is an incorporeal entity inhabiting a separate dimension from the physical body, its operations are not determined by physical law, and it possesses agent-causal power of the kind O'Connor (2000) develops. Even on these maximally permissive assumptions the constraint holds. Any conscious system, physical or ethereal, must have content in order to choose, and the very first content to populate the conscious field cannot itself be chosen, because choosing presupposes content already in place.

The argument is not causal but structural. It does not claim that the first content was caused by prior physical events, which would beg the question against dualism. It claims that choice requires prior content as a matter of necessity: a system with no content cannot perform the operation of choosing, because there is nothing for the operation to operate on. The first content arrives before choice is possible, and whatever delivers it — whether neurochemical processes or ethereal dynamics — is not directed by the conscious system's deliberate evaluation. Even if the mind existed outside the body entirely, we would have no say in what enters it first.

Hasker's (1999) emergent dualism holds that the mind is a non-physical substance that emerges from but is not identical with the brain: a "produced substance" that acquires causal powers of its own. The emergent mind, on Hasker's account, can exercise agent causation, initiating causal chains not determined by prior physical states. The

argument from initial content applies with full force, since even if the emergent mind exercises agent causation, the content on which it operates — the evaluative field that populates its deliberation — was constituted by processes it did not control. The emergent mind chooses from a menu it did not write.

Lowe's (2008) non-Cartesian substance dualism posits a non-physical self that bears psychological properties and interacts with the brain through psychophysical causation. The constraint applies unchanged: the mental states the self "owns" were constituted by processes the self did not direct, their evaluative character delivered, not authored.

5.2 The argument from ongoing delivery

Even after the initial populating, an ethereal consciousness would continue to receive content involuntarily. Whether the delivery mechanism is neurochemical — synaptic transmission, hormonal influence, neural micro-processes — or ethereal, whatever the non-physical analogue of subconscious processing might be, the relationship between the delivery system and the control centre remains the same: content is delivered to consciousness from a domain that consciousness does not direct.

A substance dualist might respond that an immaterial mind need not have a subconscious domain, since all of its processing could be conscious. The response fails for a structural reason: content-constitution and content-evaluation cannot be the same process. You cannot simultaneously be constituting your emotional response to an option and evaluating that response, because evaluation presupposes that the response is already available as an object of evaluation. The fear of flying must already be present for you to evaluate it as disproportionate, so the evaluation is a second-order operation on a first-order state that was delivered rather than chosen. The asymmetry between constitution and evaluation therefore entails a domain of processing structurally prior to the control centre, and this domain, whether physical or ethereal, is the subconscious.

5.3 The argument from perspectival structure

The perspectival asymmetry (§4) is equally substrate-independent. Any self-conscious system — any system at Level 4 or above in the categorical hierarchy (Chapter 13) — possesses the feature that generates the asymmetry: the identity of the evaluative field with the self in the present moment of choosing. An ethereal self-conscious system would experience the same identity, since in the moment of choosing it would be its evaluative field; it would be incapable of disagreeing with its current orientation. The perspectival asymmetry therefore depends on the structure of self-consciousness as such rather than on neural architecture.

5.4 The consequence

The free will problem is not a problem about the physical world; it is a problem about the structure of consciousness. Any conscious system (neurons, silicon, or immaterial substance) faces the same constraint: its evaluative field is populated by content it does not author, and its present-moment experience structurally conceals its own determination. Introducing indeterminacy does not help, because the problem is not that choices are determined but that the material of choice is delivered involuntarily — a random emotional response is no more freely chosen than a determined one. Kane's (1996) self-forming actions still operate on competing motivations whose identity, intensity, and evaluative character were delivered by the subconscious jump; indeterminacy in the resolution does not convert involuntary constitution into authorship.

6 Bounded Authorship

6.1 The concept stated

The three findings jointly produce bounded authorship. The agent is a real author at the phenomenological level — her choices flow from her evaluative hierarchy, the hierarchy expresses who she is, and the choosing is a real evaluative process rather than a confabulation — while the hierarchy from which she chooses was delivered, at the

metaphysical level, by a chain she did not initiate. A musician improvising within a given key is a real author of her melody, even though she did not choose the key: constrained rather than eliminated, her authorship real but bounded.

The boundedness applies at every level of the hierarchy, without exception. A natural challenge arises here: people appear to deliberately restructure their values through therapy, sustained reflection, religious conversion, or philosophical education.

The response is direct. An agent can initiate processes that result in value restructuring, since she can choose to enter therapy, choose to read philosophy, choose to expose herself to perspectives that challenge her existing hierarchy. But the choice to initiate the process is itself made from within the existing unchosen hierarchy: the agent enters therapy because her current hierarchy makes the prospect of change evaluatively attractive, the discomfort of her current state outweighing the discomfort of therapeutic work.

And the outcome of the restructuring process — which new values emerge, which old values are displaced, what the restructured hierarchy looks like — is not chosen but delivered by the subconscious processes that respond to the therapeutic content. The agent chose the door; she did not choose what was behind it. The authorship is real but bounded, since she initiated the process from within her unchosen hierarchy, after which the process delivered an outcome she did not select.

6.2 Frankfurt's hierarchical mesh

Frankfurt (1971) argued that freedom consists not in the absence of determination but in the alignment of first-order desires with higher-order desires. A person who wants to take a drug but wants not to want it is unfree; a person whose higher-order desires endorse her first-order desires is free. Frankfurt (1988) develops this further, arguing that what matters for freedom is not the origin of desires but the agent's wholehearted identification with them.

The convergence with bounded authorship is substantial, because both preserve practical freedom without requiring ultimate origination,

and both locate freedom in the agent's relationship to her desires rather than in the desires' causal history. The divergence comes at one point.

Bounded authorship applies the same analysis to every level of the hierarchy, since the higher-order endorsement is itself constituted by emotional responses delivered through the subconscious jump. You do not choose to endorse kindness over cruelty by an act of pure higher-order will; you endorse it because your evaluative modality responds positively to kindness and negatively to cruelty, and those responses were delivered by the same involuntary processes that delivered your first-order preferences.

Watson (1975) raised this regress worry against Frankfurt: who endorses the endorser? Bounded authorship answers that no one does, because the endorsement is itself a link in the chain, delivered rather than authored. Frankfurt correctly identifies the phenomenology of freedom, since the alignment is real and the agent experiences it as her own, while the perspectival asymmetry (§4) shows that the phenomenological accuracy does not establish metaphysical authorship. The alignment is real from the first-person present perspective; the chain that produced the alignment is visible from the third-person retrospective perspective.

Bratman (1987, 2007) extends Frankfurt's approach through planning agency: the agent's intentions structure future deliberation by organising the evaluative field around long-term commitments. Bounded authorship engages Bratman at a specific point, since the plans that structure deliberation are themselves adopted from within an unchosen evaluative hierarchy, and the sustained commitment that maintains them depends on the evaluative modality's continued oriented response to the plan's object. When the evaluative response dissolves, so that the agent no longer feels the pull of the commitment, the plan loses its structuring power regardless of the agent's cognitive intention to maintain it.

6.3 Dennett's evolved freedom

Dennett (1984, 2003) argues that the kind of free will "worth wanting" is compatible with determinism and does not require ultimate authorship of one's evaluative hierarchy. Dennett (2003) goes further, arguing that the demand for "ultimate" origination is conceptually confused, because the very idea of an agent who authors her own evaluative hierarchy from scratch is incoherent: an agent without a prior hierarchy would have no basis for choosing which hierarchy to create.

Bounded authorship shares important ground with Dennett. Both preserve reasons-responsiveness, both reject the libertarian demand for ultimate origination, and both treat the freedom worth having as real and graded. The decisive parting is over desert. Dennett holds that the determined, reasons-responsive agent is morally responsible in the full, desert-entailing sense; bounded authorship denies this, and on the question dividing compatibilism from hard incompatibilism it stands with Pereboom.

What it adds to Dennett is fourfold: the unified causal model, which Dennett's evolutionary framework does not state; the perspectival asymmetry, which explains why freedom feels real without either validating (Dennett) or debunking (Pereboom) the feeling; the ethereal mind argument's substrate-generality; and the practical/absolute distinction, which Dennett does not address because he does not engage the theological contexts in which absolute responsibility is presupposed.

The distance on the second point should be measured, because Dennett also explains why deliberation feels open, and his explanation is the one §8.5 gives first: the agent cannot predict her own decision, so the outcome is open to her (Dennett, 1984). That is an epistemic source, and the asymmetry shares it.

What the asymmetry adds is constitutive. An agent who could see the whole chain that produced her present orientation would still be unable to dissent from her resultant from outside it, since the dissent would be one more vector in the field whose resultant it opposed (§4.1). The openness therefore survives full information, as the absence of any standpoint outside the choice from which the choice could be assessed.

Dennett's openness is ignorance of an outcome; the asymmetry's is the identity of the one who would assess with the field being assessed, and the second remains when the first is removed.

6.4 Pereboom's hard incompatibilism

Pereboom (2001, 2014) argues that free will is an illusion regardless of whether determinism is true. His four-case argument is designed to show that agents in deterministic and indeterministic scenarios alike fail to meet the conditions for moral responsibility, because the causal history of their actions traces back to factors beyond their control.

On the metaphysics, bounded authorship belongs with Pereboom. It agrees that no agent has ultimate origination of her evaluative hierarchy, and, through the ethereal mind argument, that this holds whether determinism is true or false. It is, in that sense, a hard-incompatibilist position about basic-desert responsibility.

The apparent disagreement is narrower than it first looks. Pereboom does not hold that the absence of origination leaves nothing of value, since he preserves a forward-looking responsibility while denying basic desert, which is the distinction bounded authorship draws between practical and absolute responsibility. The two views converge on the substance: no basic desert, a retained forward-looking responsibility, and, in Chapter 20, restoration over retribution.

What bounded authorship adds is not a different metaphysical verdict but two contributions Pereboom's four-case argument does not supply. The substrate-independent ethereal mind argument closes the dualist and agent-causal escape routes his manipulation cases leave open. The perspectival asymmetry replaces the claim that the first-person sense of freedom is simply illusory with a more exposed one: the experience is veridical about the evaluating and silent about the authorship. The disagreement with Pereboom is therefore largely over the description of the phenomenology rather than over the metaphysics of desert.

6.5 Mele and Vargas

Mele (2006) develops the position closest to bounded authorship in the existing literature. Mele argues that the pervasive role of luck in shaping an agent's character, beliefs, and circumstances means agents cannot claim the kind of ultimate authorship libertarian free will requires. This is the broadly agnostic stance he had earlier termed "agnostic autonomism" (Mele, 1995). Vargas (2013) develops "revisionism": the folk concept of responsibility presupposes ultimate origination, which nothing satisfies, while a revised concept, grounded in the agent's capacity to respond to moral reasons, preserves everything that matters for moral practice.

The convergence with both is substantial, since all three accounts accept the reality of choice while denying ultimate authorship, and preserve practical responsibility while abandoning the folk presupposition of ultimate origination. The divergence is in the grounding.

Mele frames the limitation through luck, the contingency of the factors that shaped the agent. Bounded authorship identifies the mechanism through which luck operates: the subconscious jump and the bounded control centre, with the ethereal mind argument supplying the substrate-independence Mele's luck framework does not. Vargas gestures at the practical/absolute distinction without formally developing it or extending it to the theological contexts in which absolute responsibility is presupposed.

6.6 Levy and the role of luck

Levy (2011) argues that moral responsibility is undermined by the pervasive role of luck in shaping the agent's epistemic situation, character, and circumstances: constitutive luck. The convergence with bounded authorship is close. Where Levy emphasises the contingency of the factors that shape the agent's character, bounded authorship identifies the mechanism by which that contingency operates, the involuntary delivery of evaluative content through the subconscious jump. The convergence strengthens both accounts, since Levy provides

the philosophical analysis of luck's role while the unified model provides the explanation of why luck's role is ineliminable across any possible substrate.

This account differs from Wegner's (2002) claim that conscious will is an "illusion" generated by post-hoc confabulation. Bounded authorship disagrees: the experience of willing is a real first-person experience of evaluation, phenomenologically accurate, and what is incomplete is the authorship, since the field within which you are choosing was not itself chosen. Wegner collapses the first-person perspective into the third-person one; bounded authorship preserves both. Smilansky (2000) accepts an illusion diagnosis while arguing that the illusion should be maintained for its practical benefits; bounded authorship removes the need for such management, since what the phenomenology reports at its own scale is accurate.

6.7 Strawson's Basic Argument

The argument's nearest neighbour in the literature must be engaged by name, both to give credit and to mark the difference. Strawson (1994) argues that ultimate moral responsibility is impossible for any agent, determinism true or false. To be ultimately responsible for one's actions one would have to be responsible for one's mental nature; that would require having brought the nature about in the light of principles of choice; that would require responsibility for those principles in turn — a regress that only an impossible causa sui could halt. The Basic Argument shares the ethereal mind argument's substrate-indifference and its destination, while the mechanisms differ, a difference that matters for what each argument survives.

The two engines compared. Strawson's engine is a regress on self-creation: each act of self-formation presupposes principles not themselves chosen, iterated backwards without floor. The present engine is the constitution of the evaluative field in the present act. The first content of any deliberation cannot itself be deliberated, because deliberation operates on delivered material. The point is not that self-creation regresses but that even a single, non-iterated evaluation presupposes an evaluative endowment it did not evaluate (§5.1).

A libertarian who claims to halt the regress at self-forming actions has an answer of sorts to Strawson, since a regress can in principle be stopped at a special act. Kane's (1996) undetermined efforts at moments of inner conflict are the standard candidate. She has none to the present argument, because the self-forming act itself deliberates with responses whose identity, intensity, and evaluative character were delivered (§3.1). No indeterminacy in the resolution converts delivered material into authored material.

The constitution-based engine also yields what the regress does not: the subconscious-delivery architecture (§3) and the perspectival asymmetry (§4), which explain why the limitation is invisible from the first-person present. The two arguments are convergent and independently sufficient, so the chapter's contribution is the mechanism and the phenomenology rather than the destination. And the destination, it should be recorded, is Strawson's as much as Pereboom's.

6.8 List and Sartorio

Two further positions sit closest to the asymmetry and should be placed against it. List (2019) defends free will as real by distinguishing levels of description. Physical determinism at the fundamental level is compatible with alternative possibilities at the agential level, where the psychological states that explain action live; a science of agency conducted at that level vindicates the freedom it describes.

The kinship is real, since both views hold that two descriptions of one choice are true together and that neither reduces to the other. The divergence is plain. List's levels are both third-personal, stratifying the world's description by grain, whereas the asymmetry's perspectives are standpoints rather than levels, divided by the structure of self-consciousness. The question the levels picture leaves standing is the one bounded authorship answers. However real the agential level's possibilities, the psychological states that populate it were delivered by the subconscious jump and shaped by a hierarchy the agent did not author. An emergent level does not acquire an author by being emergent, since the delivered field is delivered at every grain.

Sartorio (2016) grounds freedom and responsibility in actual causes: an agent is responsible when her action issues from the actual sequence, sensitive to reasons, with no appeal to alternative possibilities at all. The account is the most austere compatibilism in the field; the present view can state its relation to it in one line. What Sartorio's actual-sequence freedom secures is the practical responsibility this chapter preserves; what it does not reach is the basic-desert authorship this chapter denies. The actual sequence from which the action issues includes, as its earliest links, the constitution of the very sensitivities doing the work. Reasons-sensitivity is real, causal, and gradable — the chapter's practical layer runs on it — and it is itself a delivered feature of the field. Fischer and Ravizza's (1998) guidance control, reasons-responsiveness together with the agent's ownership of the mechanism that issues in action, is the same practical layer with the ownership condition made explicit; the ownership is acquired, on their own account, through a history the agent did not author, so it secures what Sartorio's sequence secures and reaches no further.

The disagreement with both authors is accordingly small and locatable. It concerns whether anything at any level or in any sequence supplies what retribution's basic desert would need. §7 answers that question in the negative, for structural reasons no level-ascent or sequence-selection touches. The placement fixes, in passing, the chapter's relation to the debate's classic framing. Frankfurt's cases target the principle of alternative possibilities; the present argument is orthogonal to that controversy, since bounded authorship concerns sourcehood — where the hierarchy came from — rather than leeway. Its conclusion holds whichever way the alternative-possibilities dispute resolves.

6.9 The Buddhist no-self

One further neighbour predates the modern framing of the debate by two millennia and should be placed with the same care as the others. The Buddhist doctrine of no-self (anattā) denies a controller standing behind mental life. What appears as a self is a stream of dependently

arising events, the aggregates, with no owner among them; the Abhidharma analysis reduces the person to that stream (Siderits, 2003).

The convergence with the present chapter is close at three points. The delivered field: mental events arise from conditions the agent does not command, which is the subconscious jump described in a different idiom. The absence of a homunculus: §4.2 identified the "I" with the field aware of itself; the doctrine's denial is aimed at the very inner director that identification removes. And the two-level description: the tradition's distinction between conventional and ultimate truth, the person real at the first level and absent at the second. It runs parallel to the two perspectives of §4.2, each accurate within its domain (Siderits, 2003; Garfield, 2015).

The contemporary map. The tradition's bearing on moral responsibility has been worked out in the contemporary literature; the map places the chapter within it. Goodman (2002) reads the Buddhist position as a denial of desert-grounding responsibility, with compassion replacing retributive anger toward wrongdoers whose actions arise from conditions. That is a destination Chapter 20 reaches by its own routes; the convergence of independently premised paths is the kind this book treats as evidence of structure.

Siderits (1987) defends a two-level compatibilism on which agency-talk is in order at the conventional level while the ultimate level carries no self for libertarian freedom to attach to. That is the perspectival asymmetry's division of labour under other names. Repetti (2017) collects the range of readings, from hard-determinist to compatibilist. The spread itself confirms that the doctrine fixes where the debate takes place while leaving its outcome open. What the chapter adds to this neighbour is what it added to Strawson: a mechanism — the subconscious jump, the bounded control centre, the manifestation/content line of §8.6 — and a substrate-general argument that holds for any conscious system, embodied or otherwise.

The divergences are two, and each is costed elsewhere in the book. The framework retains a constituted self doing real work: the field's own self-awareness at Level 4 (Chapter 13), the bearer of practical

responsibility (§7.1). It therefore stands nearer to the conventional-person reading than to eliminativism about persons. A Buddhist who takes the ultimate level to cancel the conventional one holds a stronger claim than the asymmetry licenses, since §4.2 grades both perspectives accurate.

The tradition's contemplative claim to observe the arising of mental events is met at §8.9, where the manifestation/content line fixes what trained observation can and cannot deliver. The doctrine of dependent arising is engaged on its metaphysical side at Appendix B (§3.4), where its svabhāva-denial reading is placed beside the structural dependence of Chapter 2.

7 Moral Responsibility

7.1 Practical moral responsibility

Bounded authorship does not eliminate moral responsibility. It recalibrates it. We can and should hold agents responsible for their choices, because those choices reflect their evaluative hierarchies. When we hold someone responsible, we hold her responsible for choosing in accordance with a hierarchy that, whatever its origins, is hers.

Practical responsibility requires only that the choice reflects the agent's own evaluative hierarchy, not that she authored it from scratch. The asymmetry explains why this standard is sufficient for everyday accountability: practical responsibility operates at the phenomenological level, where the agent's authorship is real. The involuntary constitution of the hierarchy does not diminish the choice's connection to the agent: the hierarchy is hers, the orientation is hers, the choice expresses who she is.

Legal systems already operate on this basis in their treatment of excuse and mitigation, whatever desert language their sentencing retains (Chapter 20, §6). Courts hold agents responsible for choices that reflect their character while acknowledging mitigating circumstances (duress, mental illness, diminished capacity) that compromise the hierarchy's integrity. Bounded authorship provides a principled foundation for this

practice: mitigation is appropriate when the agent's evaluative hierarchy has been compromised by factors that distort its normal functioning. Mitigation is inappropriate as a blanket excuse, because the hierarchy, however it was constituted, is still the agent's, and choices flowing from it are still her choices.

Nelkin (2011) develops a "rational abilities" view of moral responsibility, on which an agent is responsible to the extent that she possessed the rational ability to act otherwise: the ability to recognise and respond to moral reasons. Nelkin's account is directly relevant. An agent with rational abilities is practically responsible, since she could have recognised and responded to the relevant reasons, while the rational abilities themselves were constituted by developmental processes she did not author. Nelkin's proportionality follows from the analysis: responsibility should be proportionate to the degree of rational control the agent exercised, and no agent exercised ultimate control over the hierarchy that constituted her rational abilities.

7.2 Absolute moral responsibility

Bounded authorship undermines a stronger form of moral responsibility: absolute moral responsibility, the claim that the agent is ultimately and completely accountable for her choices, such that infinite consequences are a proportionate response. Absolute moral responsibility presupposes that the agent had complete control over the factors that determined her choice — over the evaluative hierarchy that populated the deliberative process, and not only over the deliberative process itself.

The target of the argument. A clarification about the target is essential. Absolute moral responsibility is not a position any philosopher in the secular free will debate explicitly defends; the chapter should not be read as if it were one; what the chapter delivers against the live secular positions is the luck objection with a mechanism and a phenomenological refinement; its target is the theological tradition's ordinary libertarianism. Even libertarians such as Kane (1996) do not claim that agents create their evaluative hierarchies ex

nihilo; they claim that agents make self-forming actions at moments of inner conflict.

The target of the present argument is the theological tradition specifically: the doctrine of divine retribution, in which God holds agents eternally accountable for their choices, implicitly presupposes that the agent had ultimate authorship of the hierarchy that produced those choices. Eternal consequences are proportionate only if the agent's freedom was correspondingly ultimate. The secular free will debate operates with weaker standards of responsibility, standards bounded authorship does not challenge. What it challenges is the theological escalation from practical to absolute, which is required to justify the infinite stakes of divine reward and punishment.

The unified model makes the challenge concrete. The choice is the product of the same chain structure that governs rocks, differentiated only by the evaluative modality, its forces delivered by the subconscious jump and processed within a bounded control centre. And the ethereal mind argument shows this holds for any possible agent. No conscious system, in any possible world, made of any possible substance, ever authored its own evaluative hierarchy.

One figure in what follows must be borrowed before it is used. The omniscient judge is granted arguendo, because the tradition of divine retribution requires him, just as Chapter 20 (§6) grants omnipotence, while the book's own case against the classical conception stands (Chapters 8, 10, 17). The argument shows that even on the loan, retribution fails.

An omniscient being who understood the structural architecture of consciousness — who saw the chain from the third-person retrospective perspective with perfect clarity — would recognise that absolute responsibility is not a category that applies to any conscious agent. The infinite stakes of divine retribution, eternal punishment for wrong choices and eternal reward for right ones, are disproportionate to the structurally limited freedom with which the choices were made.

Holding an agent eternally responsible for the outputs of an unchosen hierarchy is analogous to punishing a rock for rolling

downhill — analogous at the one level the comparison licenses, ultimate authorship of the chain, and at no other. The agent, unlike the rock, has an evaluative hierarchy that is hers, which is why practical responsibility applies to her and never to the rock (§8.8). What the two share is only what absolute responsibility would need and neither has.

The rock's trajectory was determined by forces it did not author. The agent's choice was determined by forces she did not author. The forces differ in kind, gravitational versus emotional, while the chain's architecture is the same. A judge who understood this would find no desert for retribution to answer to, at any magnitude, and what remains proportionate to the structural conditions under which conscious agents make choices is the restorative response Chapter 20 (§6) develops.

8 Objections and Replies

8.1 The compatibilist objection

Objection. The perspectival asymmetry redescribes compatibilism. Compatibilists already hold that the agent is free from the practical standpoint while being causally determined from the theoretical standpoint, so the perspectival asymmetry merely relabels these two standpoints without adding philosophical content.

Reply. The reply must first correct the objection's picture of compatibilism. The sophisticated compatibilist does not claim libertarian or first-person-validated freedom. She holds, with this chapter, that the agent is fully determined, and that freedom just is the determined, reasons-responsive process. So "phenomenologically real, metaphysically determined" does not by itself distinguish the present view, since it is close to what the compatibilist already says.

The real divergence is over desert. The compatibilist holds that the determined, reasons-responsive process grounds desert-entailing responsibility; bounded authorship denies this, retaining only forward-looking practical responsibility and reserving the desert-grounding kind for the absolute responsibility it rejects (§7). On that question the view sides with the incompatibilist rather than the compatibilist, and the

premise that decides it should be named: that basic desert requires ultimate sourcehood, so an agent deserves suffering for her choice only if the self from which the choice issued was her own doing.

The premise is the incompatibilist's, the compatibilist rejects it by name, and the Premise Register carries it as a contested premise rather than a derived one. Chapter 20's case against retribution runs by two routes and only one of them uses this premise, so a reader who denies it loses the authorship route and keeps the verdict by the value route through Chapter 14. What the view then adds, and what no compatibilist account supplies, is the substrate-independent demonstration that the constraint holds under dualism and agent causation alike, together with the perspectival asymmetry's explanation — the identity of chooser and choice in the present moment — of why a determined agent necessarily experiences her determination as openness.

8.2 The libertarian objection

Objection. Agent causation rescues freedom. The agent, as a substance, can originate causal chains not determined by prior states (O'Connor, 2000; Clarke, 2003), so the ethereal mind argument addresses only event-causation rather than substance-causation.

Reply. Agent causation, even if granted, operates on material that was involuntarily constituted, so the libertarian faces a dilemma. If agent-causal power is exercised within the evaluative field already constituted by involuntary processes, then the agent is free only in the limited sense bounded authorship allows: she originates a causal chain, while the chain's content — the preferences, the evaluative orientations, the emotional responses — was delivered rather than authored.

If agent-causal power is exercised independently of the evaluative field, so that the agent can originate preferences with no basis in her prior evaluative history, then the resulting preferences are arbitrary rather than free. A preference that originates without any prior evaluative basis is a random event rather than a free choice. Clarke's (2003) integrated agent-causation theory, which requires that the agent's reasons play a role in the causation while the agent contributes an

irreducible substance-causal element, faces the same structural constraint, since the reasons that play a role are themselves evaluative orientations delivered by the subconscious jump.

8.3 The self-shaping objection

Objection. Agents reshape their evaluative hierarchies through deliberate practice, therapy, education, and habituation. This self-shaping capacity shows that the hierarchy is not entirely involuntary.

Reply. The capacity is real, and §6.1 has already given the analysis: the agent chose the door; she did not choose what was behind it. Self-shaping is agency exercised within the structural limits rather than an escape from them. The agent can modify the secondary objects in her field, altering its added-value profile (Chapter 15), while being unable to choose the primary-value responses the modifications will trigger, and every act of self-modification is initiated by a hierarchy that was itself involuntarily constituted. The bootstrapping problem remains.

8.4 The hard incompatibilist objection

Objection. If the evaluative hierarchy is involuntarily constituted at every level, why call the result "bounded authorship" rather than "no authorship"? Pereboom (2014) argues that the distinction between constrained authorship and no authorship is unmotivated.

Reply. The perspectival asymmetry provides the motivation. From the first-person present, the agent evaluates, prefers, and chooses — not a confabulation (Wegner) or an illusion (Pereboom) but an accurate report of the phenomenological condition. The hard incompatibilist collapses the first-person perspective into the third-person one; bounded authorship preserves both as accurate within their domains. The musician of §6.1 makes the point. Hard incompatibilism says that because you did not choose the key, the melody is not yours, and the asymmetry answers that the melody is yours from the perspective in which you are creating it, the key unchosen from the perspective in which the constraints are visible. Both claims are true, and neither eliminates the other.

8.5 The phenomenological objection

Objection. The experience of agonising over a moral dilemma is phenomenologically inconsistent with the claim that the outcome is determined by an involuntarily constituted hierarchy. We experience openness — the felt sense that the outcome is not yet settled.

Reply. The experience of openness is real, and the chapter explains its source. From the first-person present, the outcome is open because the agent has no access to the subconscious processes that will determine which consideration carries the most weight. The field is being constituted in real time — emotions delivered, associations activated, competing responses intensifying or fading — and the agent experiences the process as deliberation. The openness is a feature of the perspective from which the determination is invisible rather than a metaphysical feature of the choice itself.

That epistemic source is one the asymmetry shares with Dennett (1984; §6.3). The asymmetry adds a second source that does not depend on ignorance: even an agent who saw the whole chain could not stand outside her resultant to assess it, since the assessing would be a further vector in the field (§4.1), so the felt openness has a constitutive ground that full information would leave in place.

8.6 The deliberating-incompatibilist objection

Objection. The inaccessibility claim is refuted by anyone who accepts this chapter. The convinced hard incompatibilist, in the moment of choosing, knows that her current orientation was delivered by processes she did not author. She is in the first-person present; she has the metaphysical truth in mind; nothing is inaccessible to her.

Reply. What she has in mind is a content; what remains inaccessible is a manifestation (§4.2). The proposition “this orientation was delivered” is itself one more delivered content, since she does not watch her belief in the thesis arrive any more than she watches her hunger arrive; both are simply there. She knows the thesis and cannot experience its instance, because the witnessing of this preference’s production, as it occurs, is not among the field’s possible contents: a

production that appeared would be an appearance, a content, and not the producing of one. The stage can display a placard reading “the scenery is raised by machinery below”; the placard is one more piece of scenery, while the machinery is as invisible as before.

This is why her deliberation proceeds, phenomenologically, just as the libertarian’s does — a fact the account predicts and its rivals must explain away. If the sense of openness were a simple illusion, exposure should diminish it. It does not, because what exposure delivers is a content, while the openness was never a matter of contents. The persistence of the phenomenology in those who deny its metaphysical import is not a counterexample to the asymmetry; it is the asymmetry, exhibited.

8.7 The moral education objection

Objection. If the evaluative hierarchy is involuntarily constituted, moral education is pointless.

Reply. The opposite follows, since moral education is one of the primary mechanisms through which hierarchies are constituted. Education shapes hierarchies from outside the agent’s control centre, through experiences that trigger emotional responses the agent does not command; this is why it works: a child’s empathic capacity is cultivated by exposure to others’ suffering, and no act of will produces it. The unified model explains the effectiveness, because education is input into the causal chain, shaping the evaluative forces that will populate the child’s future deliberation. The involuntary constitution of the hierarchy is therefore not an obstacle to education; it is what makes education possible.

8.8 The unified model objection

Objection. The comparison between a rock and a person is reductive. Intentional behaviour involves evaluation, deliberation, and self-consciousness — features the rock’s behaviour entirely lacks. The “unified model” obscures these differences by insisting on a structural identity that ignores everything important about human agency.

Reply. The unified model does not deny the differences; it identifies them one by one. Intentional behaviour is differentiated by the evaluative modality (Chapter 14) and self-consciousness (Chapter 13), real and consequential additions that do not change the chain's architecture — the sequence of causes producing an outcome the final link did not author. The difference is in the kind of forces within the chain, physical against emotional, rather than in the chain's relationship to authorship.

The claim that the comparison is reductive mistakes a structural identity for an ontological one. The chapter does not claim that persons are rocks. It claims that both are endpoints of chains they did not initiate, and that recognising this is essential for understanding why freedom feels real while authorship is bounded.

8.9 The trained-introspection objection

Objection. The inaccessibility claim assumes an ordinary introspector. Contemplative traditions train practitioners over thousands of hours in the observation of mental events, and their reports describe watching intentions, feelings, and preferences arise: the delivery itself, witnessed. The empirical literature adds weight, since experienced meditators show measurably improved introspective access in specific domains (Fox et al., 2012), while the monitoring skills the practices cultivate are well characterised (Lutz, Slagter, Dunne, and Davidson, 2008). If any observer can watch the field being populated, the asymmetry rests on a contingent limitation of untrained minds, and its structural billing fails.

Reply. The reply grants the training and holds the line where §8.6 drew it, between a content and its manifestation. What the practised observer gains is earlier and finer access to arrivals: she detects an intention at its first flicker, notices the pull before it recruits action, and reports the sequence of deliveries with a resolution the novice lacks. Each of these achievements is an achievement about contents, since the arrival noticed sooner is still an arrival, one more delivered item in the field, and noticing it is a further delivery.

What the argument denies her is the production of an arrival as an occurrence within the field, a denial that is structural in the sense §4.2 fixed: a producing that appeared would be an appearance, so the field can contain reports of ever-earlier arrivals and never the constituting of one.

The empirical record fits this reading. The demonstrated gains concern accuracy about one's own states — image vividness, emotional tone, bodily condition — and the tradition's own vocabulary describes observing thoughts "arise and pass", which is a report about the temporal edge of contents, so the framework predicts such reports should improve with training while the delivery stays off-stage. A finding that would tell against the asymmetry has a stateable shape: an observer whose reports track the subconscious constitution of a specific preference in advance of its arrival, verified against independent measures of that constitution. The literature contains gains of the first kind and none of the second, so the asymmetry is staked on the difference remaining.

9 Conclusion

The chapter yields three results, all constructed.

The unified causal model. Intentional and unintentional behaviour share the same chain architecture — attractive and repulsive forces producing an outcome, differentiated only by the evaluative modality — with the intentional chain's forces delivered by the subconscious jump and processed within a bounded control centre. The ethereal mind argument extends this to any conscious system through the arguments from initial content, ongoing delivery, and perspectival structure.

The perspectival asymmetry. The first-person present and third-person retrospective perspectives deliver structurally divergent but simultaneously accurate reports of the same choice: freedom phenomenologically real from within, since the chooser and the choice are identical, and determination visible from without. The asymmetry is a structural consequence of self-consciousness, surviving even full theoretical knowledge of the determination, which enters the field as

one more content without making the delivery witnessed (§8.6). It explains the debate's intractability, since each side describes one perspective accurately and mistakes it for the whole.

Bounded authorship. The agent is a real author at the phenomenological level and not an author at the metaphysical level. Practical moral responsibility is preserved, since her choices reflect a hierarchy that is hers. Absolute moral responsibility is structurally undermined, since the hierarchy was involuntarily constituted, in any possible agent in any possible world, so the infinite stakes of divine retribution are disproportionate to the structurally limited authorship with which choices are made.

The chapter's position can now be placed. On the metaphysics of moral responsibility it is a hard-incompatibilist view, denying that any agent has the basic-desert authorship that absolute responsibility would require, since the ethereal mind argument shows this holds under determinism, indeterminism, and substance dualism with agent causation alike. It is therefore not compatibilism, since it denies that the determined, reasons-responsive process grounds desert, and not libertarianism, since the same constraint defeats the libertarian's escape routes.

What distinguishes it within the hard-incompatibilist family is not a different verdict but two contributions the existing positions do not provide: the substrate-independent ethereal mind argument, and the perspectival asymmetry, which corrects the standard hard-incompatibilist claim that the sense of freedom is illusory — the experience is veridical about the evaluating, though silent about the authorship. Like Pereboom, and unlike the libertarian and the retributive theologian, it preserves a forward-looking practical responsibility while denying basic desert, and Chapter 20 draws the consequence for punishment.

If morality is empathic and authorship is bounded, what follows for purpose? The agent's sense of purpose — her experience of things mattering, of life having direction and value — depends on her evaluative modality's capacity to constitute evaluative differences. But

purpose is a structural consequence of being the kind of system that can care. Chapter 19 develops the derivation.

Grades and Dependencies

1. Stakes. All three results are *constructed*, on Chapters 13–15. The chapter claims no priority on the shared ground (§1); the narrowest dispute with compatibilism is desert (§8.1), and the "absolute" responsibility denied is the theological grade only (§7.2). Two stakes are on record: the inaccessibility claim's refutation shape (§8.9) — verified reports tracking a preference's subconscious constitution *in advance* of arrival — and the omniscient judge granted arguendo (§7.2), so the retribution verdict holds even on the loan. The Premise Register carries the chain.

2. Positions placed (§6). Frankfurt: endorsement itself delivered — Watson's regress answered "no one". Dennett: shared reasons-responsiveness; parting over desert. Pereboom: same metaphysics; the correction is phenomenological. Mele/Vargas/Levy: luck given its mechanism. Wegner: first-person collapse refused. Strawson: nearest neighbour — regress engine vs constitution engine; Kane answers the first, not the second. List: levels are both third-personal; the perspectives are standpoints. Sartorio: actual-sequence freedom = the practical layer, not basic desert. Buddhist no-self (§6.9): the inner director denied with the tradition, the conventional person retained, the desert verdict shared with Goodman. Orthogonal to PAP: sourcehood, not leeway.

References

Anscombe, G. E. M. (1957). *Intention*. Blackwell.

Bratman, M. E. (1987). *Intention, plans, and practical reason*. Harvard University Press.

Bratman, M. E. (2007). *Structures of agency: Essays*. Oxford University Press.

Clarke, R. (2003). *Libertarian accounts of free will*. Oxford University Press.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. Putnam.

Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy*, 60(23), 685–700.

Dennett, D. C. (1984). *Elbow room: The varieties of free will worth wanting*. MIT Press.

Dennett, D. C. (2003). *Freedom evolves*. Viking.

Enoch, D. (2006). Agency, shmagency: Why normativity won't come from what is constitutive of action. *Philosophical Review*, 115(2), 169–198.

Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge University Press.

Fox, K. C. R., Zakarauskas, P., Dixon, M., Ellamil, M., Thompson, E., & Christoff, K. (2012). Meditation experience predicts introspective accuracy. *PLoS ONE*, 7(9), e45370.

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy*, 68(1), 5–20.

Frankfurt, H. G. (1988). *The importance of what we care about: Philosophical essays*. Cambridge University Press.

Garfield, J. L. (2015). *Engaging Buddhism: Why it matters to philosophy*. Oxford University Press.

Goodman, C. (2002). Resentment and reality: Buddhism on moral responsibility. *American Philosophical Quarterly*, 39(4), 359–372.

Hasker, W. (1999). *The emergent self*. Cornell University Press.

Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, 310(5745), 116–119.

Johansson, P., Hall, L., Sikström, S., Tärning, B., & Lind, A. (2006). How something can be said about telling more than we can know: On choice blindness and introspection. *Consciousness and Cognition*, 15(4), 673–692.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kane, R. (1996). *The significance of free will*. Oxford University Press.

Kim, J. (1998). *Mind in a physical world: An essay on the mind-body problem and mental causation*. MIT Press.

Korsgaard, C. M. (1996b). *The sources of normativity*. Cambridge University Press.

Korsgaard, C. M. (2009). *Self-constitution: Agency, identity, and integrity*. Oxford University Press.

Levy, N. (2011). *Hard luck: How luck undermines free will and moral responsibility*. Oxford University Press.

List, C. (2019). *Why free will is real*. Harvard University Press.

Lowe, E. J. (2008). *Personal agency: The metaphysics of mind and action*. Oxford University Press.

Lutz, A., Slagter, H. A., Dunne, J. D., & Davidson, R. J. (2008). Attention regulation and monitoring in meditation. *Trends in Cognitive Sciences*, 12(4), 163–169.

Mele, A. R. (1995). *Autonomous agents: From self-control to autonomy*. Oxford University Press.

Mele, A. R. (2006). *Free will and luck*. Oxford University Press.

Nagel, T. (1986). *The view from nowhere*. Oxford University Press.

Nelkin, D. K. (2011). *Making sense of freedom and responsibility*. Oxford University Press.
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231–259.
O'Connor, T. (2000). *Persons and causes: The metaphysics of free will*. Oxford University Press.
Pereboom, D. (2001). *Living without free will*. Cambridge University Press.
Pereboom, D. (2014). *Free will, agency, and meaning in life*. Oxford University Press.
Repetti, R. (Ed.). (2017). *Buddhist perspectives on free will: Agentless agency?* Routledge.
Sartorio, C. (2016). *Causation and free will*. Oxford University Press.
Schopenhauer, A. (1839/1985). *On the freedom of the will* (K. Kolenda, Trans.). Blackwell.
Siderits, M. (1987). Beyond compatibilism: A Buddhist approach to freedom and determinism. *American Philosophical Quarterly*, 24(2), 149–159.
Siderits, M. (2003). *Personal identity and Buddhist philosophy: Empty persons*. Ashgate.
Smilansky, S. (2000). *Free will and illusion*. Oxford University Press.
Strawson, G. (1994). The impossibility of moral responsibility. *Philosophical Studies*, 75(1–2), 5–24.
Strawson, P. F. (1962). Freedom and resentment. *Proceedings of the British Academy*, *48*, 187–211.
Vargas, M. (2013). *Building better beings: A theory of moral responsibility*. Oxford University Press.
Watson, G. (1975). Free agency. *Journal of Philosophy*, 72(8), 205–220.
Wegner, D. M. (2002). *The illusion of conscious will*. MIT Press.

CHAPTER 19

The Constitution of Purpose

Presupposes: the three-cell partition of mental features (Ch. 13, §2.3); the response-constitutive account of value (Ch. 14); the primary/added value taxonomy (Ch. 15); the great moral conflation (Ch. 16); the unified causal model (Ch. 18).

Establishes: the bipartite structure of purpose; its two constitutive forms; the division of the meaning-of-life impasse into a structural question answered here and a normative question referred to Chapter 16.

In brief. Does the universe care that you exist? No. Does that settle whether your life has meaning? Also no — the inference from indifference to meaninglessness is a non sequitur. Purpose has two parts: an object (the target) and a sense (the felt pull toward it). A map of a city is a representation; a desire to visit the city is a purpose. Two constitutive forms, aesthetic and empathic; standards borrow their pull from these. The nihilist reporting that nothing matters is the blind person concluding that colour does not exist: the modality has dimmed, not the world. The meaning-of-life impasse divides: subjectivists describe the sense, objectivists the object, hybrids the conjunction the architecture entails — and the objectivist's further question, whether a purpose is worth having, passes to Chapter 16. The response is involuntary; the world is not of our making; but the conjunction is real. It is ours.

---

1 Introduction

Does the universe care that you exist? No. Does that settle the question of whether your life has meaning? Also no, and the inference from cosmic indifference to personal meaninglessness is one of the most persistent non sequiturs in the history of philosophy. The universe does not need to care for your life to matter. What matters is that you care, since your evaluative modality constitutes oriented differences in response to the world. Purpose is therefore not something added to consciousness from outside; it is constituted by the evaluative architecture the preceding chapters have developed.

The contemporary meaning-of-life debate is structured by a familiar impasse.

Naturalists hold that the universe is purposeless and that meaning is a projection of human psychology onto indifferent matter (Nagel, 1971; Camus, 1942/1955). Theists hold that purpose is a divine endowment (Craig, 2008; Cottingham, 2003). Hybrid theorists attempt to preserve both elements (Wolf, 2010). The present chapter divides the impasse: its structural half is resolved by showing that the three positions are partial descriptions of a single phenomenon, and its normative half is handed to Chapter 16. The phenomenon is one whose internal structure parallels the bipartite architecture already established for aesthetic experience, empathic experience, and certainty.

The chapter develops the object/sense distinction for purpose (§2), then integrates purpose with the hierarchy of value and the dominant purposes that shape life direction (§3). It identifies two constitutive forms of purposive orientation, aesthetic and empathic, and shows that orthotic purpose is derivative (§4). It develops the phenomenology and structural diagnosis of purpose-loss and nihilism (§5), divides the meaning-of-life debate through sustained engagement with Wolf, Metz, Frankfurt, de Beauvoir, Nagel, and Setiya (§6), and addresses objections (§7).

2 The Object and Sense of Purpose

2.1 The bipartite structure

Purpose, like every other evaluative category in the analysis, has a bipartite structure: an object and a sense. The object of purpose is the target toward which the agent is oriented — a career, a family, a cause, a creative project, a relationship. The sense of purpose is the emotional drive or orientation toward that target — the felt pull, the engagement, the evaluative response that makes the target matter rather than merely being represented.

The distinction is not merely terminological. It names two different components whose conjunction constitutes purpose and whose separation explains its absence. A person who identifies a target without any emotional engagement has a representation, not a purpose. A map

of a city is a representation; a desire to visit the city is a purpose. The difference is the presence or absence of the sense — the evaluative modality's oriented response. Conversely, a person who experiences a diffuse emotional drive without identifying a target has a mood, not a purpose. The restlessness of someone who feels she should be doing something without knowing what is the sense of purpose seeking its object.

The distinction can be stated in a sentence: the emotion we experience corresponds to the sense of purpose, while the thing we are drawn to — a family, a cause, a career — is its object. It yields a compact definition: a purpose is a desired prospect or predicted outcome, with the outcome supplying the object and the desire supplying the sense; that a purpose ordinarily issues in action is a consequence of the sense.

2.2 Structural parallels

The object/sense distinction for purpose is not an isolated observation, since it instantiates the bipartite architecture that runs through the entire evaluative framework. Aesthetic experience has an object of beauty, the sunset or the melody, and a sense of beauty, the modality's oriented response. Empathic experience has an object, the other's suffering or the friend's joy, and a sense, the oriented response to the other's state.

In each case both components are structurally required. Strip away the sense and the object remains a neutral stimulus. Strip away the object and the sense has nothing to respond to, since a detection without evaluative response is a perception rather than an experience of beauty or of empathy.

Certainty exhibits the same bipartite structure, with one refinement of its own. The object of certainty is the claim itself, the content the conviction is about, while the sense of certainty is the felt conviction, Chapter 17's emotional-pull axis. What Chapter 17's logical axis adds is not a further component but a graded status the object bears: its denial can be truly held or it cannot.

Certainty is therefore the partial instance of the pattern — object and sense like the others, plus an accuracy instrument the others lack. Rational certainty obtains when the sense tracks the object's logical status; irrational certainty obtains when the pull toward closure overrides an open status (Chapter 17, §2.3). The aesthetic and empathic cases have no analogue of this grading, which is why dogma is an epistemic pathology while there is no corresponding pathology of finding sunsets beautiful.

Purpose completes the pattern, which is an inherited partition rather than an induction over cases. It is the two axes of the mental (Chapter 13, §2.3): every feature of a mental state is a content or the orientation of one, an exhaustiveness derived rather than observed, and secure on the derived tier even for a reader who rejects the placement of feltness.

The object of purpose is therefore the content the response is directed at, the sense is the orientation; there is no third component for a purpose to have, because there is no third kind of mental feature for it to be. The survey above is display rather than proof, since aesthetic experience, empathic experience, and certainty exhibit the structure because everything mental does. Purpose is not an exception to this architecture but its most comprehensive instantiation, because the object of purpose can be anything in the agent's evaluative field — any object, activity, relationship, or goal — while the sense of purpose is the evaluative modality's sustained oriented engagement with that object.

2.3 The derivation

The derivation proceeds through four claims established in earlier chapters. All content requires distinction (Chapter 2). Consciousness is constitutively dependent on detected difference (Chapter 13). The evaluative modality constitutes evaluative difference through oriented response to non-evaluative input (Chapter 14). And the evaluative field is involuntarily constituted by the subconscious jump (Chapter 18).

From these four claims, purpose follows as a structural consequence. A conscious system with a functioning evaluative modality, operating within a field of differences, is a system for which

things matter. Its evaluative modality constitutes oriented differences (pulls toward some possibilities and away from others) and this constitution is the core of purpose. Purpose obtains when the system identifies an object and the evaluative modality constitutes an oriented response — the sense — toward that object.

The derivation's yielding this result is itself notable, because purpose might have seemed an axiological supplement added from outside. That it follows as a structural consequence matters not because the conclusion is exotic — that creatures with evaluative modalities find things mattering is familiar — but because a derivation, unlike an observation, carries commitments, and the commitments are what the rest of the chapter must defend.

Purpose, on this account, is not added to consciousness from outside. Wherever the conjunction of consciousness, the evaluative modality, and a field of non-evaluative differences obtains, purpose obtains. Wherever it does not (at zero complexity, where no differences are available; in a system without an evaluative modality; in a system whose evaluative modality has been silenced) purpose is structurally absent. Purpose is present wherever the evaluative modality is engaged with real differences. It is absent at the core of a star, in the emptiness of space, and in any system whose evaluative modality has ceased to operate.

2.4 What purpose is not

Four things purpose is not can now be stated. Purpose is not a cosmic telos, since the account does not claim that the universe has a purpose or that reality is structured toward an end. Purpose is not a divine endowment, since the first section of this book established that the grounding terminus is unintelligent and that creation ex nihilo is structurally incoherent. Purpose is not a subjective projection, since the evaluative differences the modality constitutes are response-dependently real rather than manufactured from nothing. And purpose is not a brute psychological fact, since the structural account explains why things matter to conscious systems by connecting purpose to the analysis's foundational results.

3 Purpose and the Hierarchy of Value

3.1 How the hierarchy generates purpose

The derivation establishes that purpose is a structural consequence of evaluative engagement. But purpose does not float free of the evaluative architecture developed in Chapter 15. Purpose is anchored in the hierarchy of value — the ordered ranking of objects by the intensity and orientation of the evaluative responses they elicit. The hierarchy is not merely a context in which purpose occurs; it is the mechanism through which purpose is constituted, sustained, and transformed.

A purpose is a sustained evaluative orientation maintained across time through the hierarchy of value. A passing desire is a momentary orientation — a transient evaluative response that fades with the situation that triggered it. The difference is temporal persistence sustained by the hierarchy: a purpose occupies a stable position in the hierarchy, anchored by primary and added value that resist displacement. A desire occupies a transient position, anchored by primary value alone, without the web of derived and consequential associations that sustain a purpose across changing circumstances.

Purpose admits of degrees along a single dimension: the intensity and stability of the evaluative orientation. A person whose evaluative modality responds to a target with strong, sustained, and richly associated attraction has a deep purpose. A person whose response is weaker, less stable, or more thinly associated has a shallower purpose. The degree tracks the evaluative modality's actual response rather than the subject's reflective endorsement of it: a person can articulate a purpose she does not deeply feel, and feel a purpose she cannot articulate.

Purposes conflict when two or more targets occupy positions in the hierarchy that generate incompatible behavioural orientations. A person who is oriented toward both career success and family presence faces a conflict of purposes when the two require incompatible allocations of time.

The conflict is resolved, if it is resolved, by the forced-choice diagnostic (Chapter 15): under the pressure of compelled selection, the evaluative modality produces responses of different intensities to the competing targets, and the stronger response determines the choice. The resolution may surprise the agent, since a person who believed she valued career above family may discover, in the moment of choice, that her evaluative modality responds more intensely to the prospect of her child's distress than to the prospect of professional advancement. The hierarchy reveals itself only when the choice is forced.

3.2 Dominant purpose and life direction

In any hierarchy of value, the object associated with the strongest emotional response, positive or negative, is by definition the most valuable or dominant. What the preceding chapters have not developed is the consequence of this for purpose. Whatever sits at the top of a person's hierarchy becomes her dominant purpose: the orientation that shapes not just what she chooses but who she is. Dominant purpose is the hierarchy's apex viewed from the perspective of action and identity.

Each of us lives by a unique hierarchy of values, so our decisions and life directions are shaped by factors that are dominant to us without being shared or prioritised by others. Two documented lives display the mechanism at full evidential strength; a schematic survey completes the range.

Kalanithi. Paul Kalanithi's memoir records the mechanism under the harshest available conditions: a neurosurgeon in his mid-thirties, diagnosed with metastatic lung cancer, watching his evaluative hierarchy reorganise itself under the diagnosis (Kalanithi, 2016). Two orientations dominate the record. The first is self-empathic — the drive toward survival and, as survival narrowed, toward time that mattered, the modality responding to his own suffering with an oriented pull toward its relief. The second is empathic and outward — the decision to father a child he knew he would not raise, and to write for readers he would never meet: sustained orientations toward the emotional lives of others.

The memoir illustrates the diagnostic rather than instancing it, since a single life is not the insulated repetition Chapter 15 (§4.1) requires before a compelled choice can be said to disclose an ordering rather than constitute one. It is cited for the shape of the mechanism rather than as evidence of it. The surgical career he had spent a decade building was displaced within months; the displacement arrived as a discovery about his own hierarchy rather than a decision made over it.

Frankl. Viktor Frankl's account of the camps is the classical narrative case; it is striking that the founding text of the meaning-of-life literature is a report of dominant purpose under conditions engineered to destroy it (Frankl, 1946/2006). Frankl records prisoners whose orientation toward a specific object — a spouse who might still be alive, a manuscript to reconstruct, a child waiting — organised perception, endurance, and conduct after every ordinary support had been removed. His own sustaining object was the confiscated book he intended to rewrite.

In the present vocabulary the observation is apt. The object of purpose was in each case a target in the world; the sense of purpose was the empathic or self-empathic orientation toward it; and the intensity of that orientation, delivered involuntarily, marked the difference Frankl reports in his own case and in the cases he could observe — a report from a survivor about survivors, which is why it is cited for the shape of the mechanism and licenses no verdict about those who did not survive.

The caveat entered for Kalanithi applies here with more force, since a memoir written by a survivor, about survival, is the least controlled evidence a causal claim of this shape could rest on. Frankl built a therapeutic practice on the observation; the structural account explains it. Whatever sits at the top of the hierarchy shapes conduct and identity because it commands the strongest oriented response.

The mechanism is indifferent to the content it operates on. A dominant purpose may be anchored in gratitude, in the anticipated beauty of achievement, in hope of recovery, in solidarity with the oppressed, or in the promise of salvation. The hierarchy's dominant value supplies the object; the modality's oriented response constitutes

the sense; the two together shape identity and life direction. The mechanism is the same in every case; only the content varies.

3.3 The evaluative hierarchy is involuntarily constituted

The evaluative hierarchy is involuntarily constituted. Purpose, while real, is therefore not self-authored in the ultimate sense. The agent did not choose which evaluative differences her modality constitutes most vividly, which targets attract her most strongly, or which forms of purposive orientation dominate her life. Kalanithi did not choose the force of his survival drive or the pull toward the daughter he would not raise; Frankl's prisoners did not choose which remembered face or unfinished task would carry the strongest charge. Purpose is real and purpose is theirs; it is a product of bounded authorship (Chapter 18). The implication for divine punishment — that holding agents to absolute account for purposes they did not fully author is disproportionate — is developed in Chapter 20.

4 Two Constitutive Forms of Purpose

4.1 The aesthetic and empathic constitution

The three-goodness taxonomy (Chapter 14) identifies two constitutive forms of evaluative engagement, aesthetic and empathic, alongside a third form of goodness, orthotic, that is a cognitive operation rather than an evaluative modality. A natural expectation is that purpose would exhibit the same three forms. It does, though with one asymmetry that the taxonomy itself predicts.

Aesthetic purpose obtains when the evaluative modality responds to sensory or representational content, so that the system is oriented toward beauty, harmony, elegance, or the pleasures of perception. A painter oriented toward creating a beautiful image, a musician toward producing a satisfying sound, a hiker toward experiencing a landscape: each is engaged in aesthetic purpose. A life rich in aesthetic purpose is one in which the evaluative modality is frequently and vividly engaged with the sensory world.

Empathic purpose obtains when the evaluative modality responds to the emotional states of others, or of oneself in the self-empathic case. A parent comforting a distressed child at three in the morning — the exhausted walk to the nursery, the warm bottle, the slow rocking — is engaged in empathic purpose, the evaluative modality responding not to the beauty of the task, since there is little, but to the child's distress and the parent's need to relieve it. A doctor relieving pain, a friend supporting someone through grief, a scientist who spends decades pursuing a cure for a disease affecting children she will never meet: each is oriented by empathic purpose.

Self-empathic purpose obtains when the modality responds to one's own emotional states — the desire to flourish, to feel well, to escape suffering. Kalanithi's survival drive is self-empathic purpose: the evaluative modality responding to his own suffering with an oriented pull toward recovery and, when recovery closed, toward the time that remained (§3.2).

This form, empathic purpose in both its other-directed and self-directed varieties, is the most directly connected to morality and, for many people, the form most resilient in the face of mortality. The knowledge that one's empathic engagement made a real difference to another person's emotional life is a form of significance that requires neither cosmic endorsement nor infinite duration.

4.2 Orthotic purpose is derivative

Orthotic purpose (the orientation toward conformity with a standard) is real as a descriptive category but is not constitutively purposive. This parallels Chapter 16's finding that morality is constitutively empathic and orthotic assessment is derivative: the orthotic standard codifies prior evaluative orientations rather than generating new ones.

The argument is straightforward; it is not new, since it is the grip argument of Chapter 16 (§3.1) in the key of purpose. A standard engages the evaluative modality or moves nothing, so what held for moral requirement holds for orientation toward any standard whatever.

What follows restates the point in this chapter's materials, as illustration rather than fresh proof. Purpose requires the sense (§2), while orthotic assessment does not constitutively require the modality, since a spell-checker assesses without emotion. The conformity-checking operation can therefore proceed without any oriented emotional response, and without the sense there is only a representation of a standard.

What, then, is "orthotic purpose"? A student oriented toward passing an examination, an athlete toward breaking a record, a craftsperson toward producing a flawless product: these feel purposive. But the purposive element is not the conformity-checking; it is the emotional response to conformity or non-conformity. The student feels anxiety about failure and relief at success — these are self-empathic responses (the evaluative modality responding to her own prospective emotional states). The athlete feels the pull of achievement: an aesthetic response (the beauty of the completed performance) or a self-empathic response (the desire to feel the satisfaction of triumph). The craftsperson feels the satisfaction of precision: an aesthetic response to the beauty of well-executed work.

Strip away the emotional response and the orthotic operation continues (the spell-checker still checks) but purpose vanishes. The student who checks her exam answers mechanically, with no emotional engagement, has no purpose; she has a procedure. The orthotic standard provides the object of purpose; the aesthetic or empathic response provides the sense. Without the sense, the object is inert.

The three forms therefore reduce to two constitutive forms, aesthetic and empathic including self-empathic, with orthotic purpose the application of these to orthotic targets. The taxonomy is preserved as a descriptive tool, while the purposive element is always aesthetic or empathic: the orthotic element provides the target rather than the purpose.

4.3 Living in others' minds

Human beings, among all objects of purpose, stand apart for the intensity and resilience of the responses they elicit; the account

identifies the mechanism, which is the one §4.2 fixed: another mind is the only object of purpose that is itself a register, so an orientation toward it meets a response; the object is renewed by the relationship — which is what the intensity and the resilience both come to. To live in the welcoming mind of another is empathic purpose in its most intimate form. The object of purpose is the other's awareness of you; the sense is the empathic pull toward being known, understood, and valued by that consciousness.

The object admits every configuration the social world supplies: concentrated in one person who knows you deeply, held within a family or a few friends, dispersed broadly as fame, lodged in the thoughts of someone already dead, or, for the believer, in the mind of a God who, on Chapter 24's proposal, registers without having authored. Scheffler's (2013) scenarios make the object visible by subtraction: told that humanity would end shortly after their own deaths, most people find their projects drained of point, which shows that the awareness of others, present and to come, was among their objects all along. The resilience follows from the structure. An awareness of you is an object that ongoing relationship continually renews, so the purpose it anchors is atelic in Setiya's sense (§6.3), consumed by no achievement.

The same structure explains the characteristic losses. Where the object is another's awareness, the purpose is exposed to that awareness's withdrawal: death, rejection, the erosion of memory. The loss of meaning such withdrawal produces is the collapse of an object, with the sense left seeking what it can no longer engage.

It explains, too, why love is valued as it is. Mutual awareness is the configuration in which two such objects sustain each other, granting significance under physical isolation: the soldier at war carried by the belief that he is remembered. And it explains why aesthetic engagement deepens when shared, since the empathic amplification of the aesthetic is the two constitutive forms arriving together, each supplying the other's object.

Where the argument stands: purpose has been given its bipartite structure and derived from the two axes of the mental (§2); placed

within the hierarchy of value as dominant purpose (§3); divided into its two constitutive forms (§4). What follows is what happens when the structure fails.

5 The Absence of Purpose

5.1 The phenomenology of purpose-loss

The phenomenological description that follows is not decorative; it is the evaluative architecture in action. Each sentence describes the evaluative modality's progressive dimming: the sense of purpose weakening, the object of purpose losing its emotional charge, the hierarchy of value flattening toward indifference. The taxonomy that follows is the formalisation of what this description captures from the inside.

The objects that anchor purpose vary across lives: the small textures of a day, a long ambition, wealth, companionship, art, inquiry, the natural world, the built one. The variation belongs to the object, while the mechanism is constant. A person in love shows the mechanism at full amplitude, since the modality responds to nearly everything, and a grey day carries a charge it would otherwise lack.

The mechanism at low amplitude shows the same structure from the other side. A person cut off from her future finds few responses strong enough to sustain the report that her life has purpose. As the responses dull, the smile registers as less happy and the joke as less funny; when positive response fails widely enough, the field flattens and its objects present as worthless.

At that point the sense of purpose may no longer be renewed from within, because thought after thought arrives without a positive charge attached. Renewal then often comes from outside — a friend, a family member, a stranger's kindness — supplying the field with an object vivid enough to re-engage the modality.

5.2 Three structurally distinct cases

If purpose is constituted by the evaluative modality's oriented response, then the absence of purpose admits of a sharp structural diagnosis. But the absence is not uniform. Three structurally distinct cases must be distinguished.

Case 1: Evaluative silence. The evaluative modality falls silent — true numbness, no positive or negative feelings. There is no purpose because no evaluative differences are being constituted. The person does not move toward or away from anything. She does not declare "nothing matters"; she simply does not engage. This is the structural zero of purpose: evaluative equalisation, the analogue of blindness. Just as the blind field is not darkness but the absence of the visual modality's operation, evaluative silence is not sadness but the absence of the evaluative modality's operation.

Case 2: Negative engagement. The evaluative modality is still operating while constituting only negative evaluative differences: pain, despair, anguish, repulsion. There is still purpose, though a purpose that on reflection is unwanted. The person is moved, since the negative feeling pushes or pulls. She has meaning in the formal sense; she does not want the meaning she has. A person overwhelmed by grief is purposive, since she is oriented toward the cessation of pain, while the purpose is one she would not choose if she could.

The clinical significance of the Case 1/Case 2 distinction deserves emphasis. Severe depression frequently involves the dominance of negative emotional response rather than its absence, and the person's report, "nothing matters", misdescribes her condition as Case 1 when it is structurally Case 2. She reports evaluative silence when what she is experiencing is negative engagement so pervasive that positive engagement has been displaced. The structural diagnosis matters because the interventions differ: Case 1 (evaluative silence, as in certain neurological conditions) requires restoring the modality's operation; Case 2 (negative engagement misreported as silence) requires rebalancing the modality's orientation, not restarting it.

Case 3: The purposive act of self-destruction. The person who ends her own life to escape unbearable suffering is performing a purposive act — oriented toward the cessation of suffering, which is an evaluative target constituted by the modality's negative response to her current state. Purpose is present in the act: the modality's last operation orients the agent toward the elimination of the conditions for its own operation.

The practical consequence follows the pattern of the two cases before it. The orientation is real, and the field constituting its object is the field Case 2 describes: positive engagement displaced, so what the modality has to weigh has narrowed to the suffering and its ending.

That narrowing is a state of the evaluative field rather than a conclusion the person has reasoned her way to, and a field in that state is altered as Chapter 15 (§4.4) says fields are altered — by enrichment rather than by argument. Section 5.1 has already located where the enrichment comes from once it can no longer be generated from within: another person, supplying an object vivid enough to re-engage the modality. Classifying the act as purposive therefore settles nothing about whether the orientation is to be honoured. What it registers is the constriction of a field, where the constriction is what admits of change.

The three-case analysis prevents a glib claim that meaning is always present as long as someone is alive. In Case 1, purpose is absent. In Case 2, purpose is present and unwanted. In Case 3, purpose is present in the act itself.

5.3 Engagements: Ratcliffe and Smuts

Ratcliffe (2015) provides the most detailed phenomenological analysis of how depression transforms experience: not primarily a mood but a transformation of existential feeling, the background sense of being situated in a world that matters, offers possibilities, solicits engagement, so that the world does not merely seem worse but closed.

What he identifies as the alteration of existential feeling is, in the analysis's terms, the diminishment of the evaluative modality's capacity to constitute oriented differences: the world seems closed because the

modality that would constitute its openness has fallen silent or become dominated by negative orientation. Ratcliffe's analysis therefore confirms the Case 1/Case 2 distinction, since some depressive states involve evaluative silence, where the world offers nothing, and others pervasive negative engagement, where the world offers only suffering.

Smuts (2013) identifies a persistent gap in the good-life literature, since theories of well-being do not adequately distinguish a life that is going well from a life that has meaning. The object/sense distinction fills it: well-being concerns the modality's overall balance of positive and negative orientation, the hedonic balance of the sense, while meaning concerns its engagement with real non-evaluative differences, the richness of the object.

Case 2 nihilism illustrates the distinction, since a person in pervasive negative engagement has poor well-being without being without purpose, the modality still constituting oriented differences directed toward objects. Well-being and meaning are therefore structurally distinct dimensions of evaluative life, for which the object/sense distinction provides the vocabulary for stating the difference plainly.

5.4 Nihilism as evaluative diminishment

A distinction the framework makes available must be drawn first. The nihilism described above is experiential — the lived collapse of the sense that anything matters — and to that condition the present account speaks directly, since it is a report of evaluative diminishment rather than a discovery about the world.

The analogy with blindness is strict. A person who loses her sight has not discovered that colour does not exist; she has lost the modality whose response constitutes chromatic differences. Similarly, a person whose evaluative modality is diminished has not discovered that nothing matters; she has lost the modality's capacity to constitute evaluative differences, the capacity to generate the sense of purpose. The nihilist's report, "nothing matters", is therefore an accurate

description of her current phenomenology and an inaccurate generalisation about the world.

Where nihilism is evaluative diminishment, the appropriate response is not philosophical argument, which targets beliefs, but modality-restoration, which targets the evaluative system's capacity to constitute the sense of purpose. This is consistent with clinical practice, where treatments for depression and anhedonia target the substrate of emotional response rather than the arguments for or against meaning.

The philosophical error theorist is a different case, and must not be diagnosed in this way. He advances a claim — that nothing matters stance-independently — rather than reporting a deficit, and his evaluative modality may be in perfect order. The blindness analogy does not refute him, and invoked against him it would beg the question, presupposing the response-dependent reality of evaluative differences that he denies. He is met on the merits instead, those of Chapter 14, on whether the secondary-quality model of value can be sustained. To treat his position as a condition rather than a claim would be the very error this book diagnoses in dogma (Chapter 17): mistaking a philosophical disagreement for a psychological one.

6 The Meaning-of-Life Debate Divided

6.1 The impasse

Subjective theories (R. Taylor, 1970; Frankfurt, 2004) hold that life has meaning when the subject cares about what she does. Objective theories (Metz, 2013; Wielenberg, 2005) hold that meaning requires engagement with activities that are worthwhile. Hybrid theories (Wolf, 2010) combine both. The impasse arises because each captures something real, while none explains why both elements are needed or how they are related. The object/sense distinction developed in §2 provides what the existing literature lacks: an explanation of why both components are required and how they are related.

6.2 The division

The subjective theorist is right that purpose requires the sense, the evaluative modality's oriented engagement. A person who performs objectively worthwhile activities without any emotional response does not experience purpose: she has an object without a sense. The objective theorist is right that purpose requires real differences in the object, the non-evaluative input the modality responds to. Some activities involve richer non-evaluative features than others, so the richness of the object affects the depth of the purpose constituted.

The hybrid theorist is right that both elements are needed, while treating the conjunction as a brute requirement. This account explains it. The sense of purpose is constituted by the modality's oriented response to the object of purpose, so both the modality's engagement (sense) and the input's real differences (object) are structurally required. The conjunction is entailed by the bipartite architecture rather than stipulated.

A clarification is needed about what the division settles and what it leaves open. The objective theorist's condition was not merely that the object contain real differences but that it be worthwhile. Rendering it as "real non-evaluative differences" captures the structural requirement — that purpose needs a real object — while setting the worthiness question aside. That question does not vanish, because it belongs elsewhere.

A person richly and stably engaged with an atrocious project, a life organised around cruelty pursued with ingenuity and devotion, meets the structural conditions for purpose: an object, and a sustained oriented response to it. On this account she has purpose, even deep purpose. What she lacks is goodness rather than purpose. The empathic account of morality (Chapter 16) delivers that verdict: a judgement about the worth of the response rather than about whether it constitutes a purpose.

What is resolved is therefore the structural impasse, why both object and sense are required for purpose to exist, while the normative question the objectivist also presses remains open here. Whether a purpose is worthwhile is a moral question, answered in Chapter 16; whether something is a purpose at all is the structural question answered here.

Keeping the two apart is what the account adds. The objectivist is answered on the structural question and declined on the normative one: where a purpose survives full registration, the account withholds the verdict of meaninglessness Wolf's condition was built to deliver; that is a residue conceded rather than a dispute won.

6.3 Wolf, Metz, Frankfurt, and de Beauvoir

Wolf's (2010) formulation, that meaning arises from "active engagement in projects of worth", comes closest to the present account. "Active engagement" captures the sense; "projects of worth" captures the object. This account agrees with the formula but provides what Wolf does not: an explanation of why the conjunction is necessary. Wolf treats it as an intuitive requirement. This analysis derives it from the response-constitutive architecture: the sense is constituted by the modality's response to the object, so both are required. The conjunction is not intuited but entailed.

The grass-counter. Wolf's formula carries a commitment the confirmation must not skate over, since it is the whole point of her objectivity condition. Wolf introduces "of worth" to exclude cases in which the engagement is real while its object is not: the man who lives for counting blades of grass, or who is fulfilled by memorising phone books, engages actively and is subjectively gripped. Yet his life does not thereby become meaningful. If purpose is constituted by the modality's oriented response, the worry runs, then the grass-counter has purpose in full, so the account collapses the distinction Wolf's condition was built to preserve.

Response-constitution is not response-infallibility. The reply is that response-constitution does not entail response-infallibility, and the resources to mark the difference are already in hand. An evaluative response is directed, since it registers the features of its object accurately or misreads them (Chapter 14, §6.4).

The grass-counter's orientation is therefore not validated by being felt. It is a response whose object bears none of the evaluative differences the response imputes to it — a misregistration, the purposive

analogue of the adaptive preference (Chapter 15, §6) and the manufactured added value. What Wolf records as a failure of objective worth, the analysis records as a failure of fit: the sense is real, while it answers to nothing in its object. A purpose whose object cannot sustain the significance the sense assigns it is criticisable on the account's own terms. The distinction Wolf draws by positing a stance-independent order of worth, the analysis draws within the response-constitutive architecture, by the directedness of the response, so Wolf's condition is replaced by an informed-attitude criterion that does its work without the objectivism.

This also fixes the account's exposure. It does not follow that every widely disvalued purpose is a misregistration, since the criterion is fit to the object's actual evaluative differences rather than conformity to consensus. Where an object bears the differences a purpose imputes to it, the purpose stands however unusual it looks. What is excluded is only the purpose whose object bears no such differences at all, which is the case Wolf's condition was meant to exclude, and no wider.

One step in this reply must be closed against a circle it invites. On a response-constitutive account, "the object's actual evaluative differences" cannot mean differences constituted independently of all response, since there are none, and if it means differences constituted by the grass-counter's own response, every purpose fits its object trivially.

The criterion is counterfactual and diagnostic — the informed-registration anchor of Chapter 14 (§6.5), of a piece with the tests for adaptive preference and manufactured added value (Chapter 15, §§6.6, 7.2). A purpose fits its object when the orientation survives the agent's own accurate and complete registration of the object's non-evaluative differences, so that fuller acquaintance with what the object actually contains sustains the response rather than dissolving it. The differences that do the criticising are non-evaluative and observer-available; the response that does the sustaining or dissolving is the agent's own, corrected only for accuracy and completeness, under the forced-choice conditions Chapter 15 specifies. For empathic purpose the anchor is firmer still, since it is the other's actual felt condition, observer-independent in the sense of Chapter 14 (§6.4).

The grass-counter is therefore criticisable not because grass lacks differences, since blades differ, but because the significance his sense assigns is one his own modality, brought into complete contact with those differences and with the rest of his evaluative field, does not sustain. The orientation lives on inattention, as the adaptive preference lives on foreclosed alternatives.

The account concedes the residue openly: a grass-counter whose orientation survives full registration and complete comparison has a purpose the analysis will not condemn. The outcome is one Wolf's objectivism forbids and the present account accepts, as the cost of drawing the objectivity condition inside the architecture rather than above it. Refuse constitution and keep Wolf's objectivism, and no residue is owed here, since the criterion then applied is the one the fit criterion approximates; the residue is borne by those who draw the condition inside.

Metz's (2013) comprehensive survey identifies a cluster of conditions — creativity, moral goodness, the exercise of rational capacities — that track meaningfulness across cases, and the analysis explains the cluster. Each involves rich non-evaluative differences, the objects, to which an engaged modality responds, the sense: the achieved against the not-yet-achieved, the empathic deployment, the solved against the unsolved. The cluster is therefore not a list of independent meaning-makers but a set of activities whose objects are particularly rich in the differences that engage the sense with particular intensity.

Telic and atelic. Setiya (2017) identifies a structural source of dissatisfaction the meaning-of-life literature has overlooked: the difference between telic activities, oriented toward a terminal goal that, once achieved, eliminates the activity, and atelic activities, oriented toward ongoing engagement not exhausted by any particular achievement. The midlife crisis, on Setiya's diagnosis, is the recognition that one's life has been organised around telic purposes — career goals, financial targets, milestones — that are either achieved, and therefore no longer purposive, or abandoned, and therefore experienced as failure.

The present account gives Setiya's diagnosis its mechanism through the object/sense distinction. Telic purposes are orientations whose object is consumed by achievement, since the contrast between goal and absence disappears once the goal is reached. Atelic purposes are orientations whose object is renewed by ongoing engagement, since the contrast between engaged and disengaged is reconstituted at every moment. The analysis therefore predicts what Setiya observes: a life rich in atelic purpose — ongoing aesthetic, empathic, and intellectual engagement — is more resilient against existential dissatisfaction than a life structured exclusively around telic achievement.

Frankfurt's (2004) account of caring as the source of importance converges on the central point: what matters to the agent is constituted by what she cares about rather than by an external authority's endorsement. What the structural account adds is the location, since caring is not a brute volitional fact but the sense of purpose, the modality operating on an object, within a hierarchy involuntarily constituted by the subconscious jump. Frankfurt identifies what purpose consists in; the analysis identifies why caring has the bipartite structure it does and why the agent's authorship of what she cares about is structurally bounded.

De Beauvoir (1947/1976) argues that meaning is neither given by the situation, as objective determinism holds, nor created ex nihilo by the will, as Sartrean radical freedom holds. It is, on her account, constituted by the agent's engaged response to a situation she did not choose and within which her choices are real. The parallel is close: the object of purpose is a field of differences the agent did not populate, and the sense is the modality's oriented response to that field. Her situated freedom is bounded authorship under another name, and her insistence that meaning requires both the situation, the object, and the engagement, the sense, anticipates the dissolution of the impasse.

Two further positions complete the map. Kauppinen (2012) argues that meaning is a property of temporally extended wholes, so a life's meaningfulness depends on its narrative shape and the same episode contributes differently according to the arc it belongs to. The account absorbs the insight without the holism.

Narrative shape is a real feature of the object, since a life's trajectory, its progressions and redemptions, are differences the modality registers, entering the composite as added value across time (Chapter 15, §3). That is why an achievement embedded in an ascent feels different from the same achievement stranded. What the bipartite structure denies is that the sense waits on the whole, because the orientation operates now, on an object that includes the shape so far and the shape projected. A view on which no episode's significance is settled until the story ends therefore mistakes a feature of the object for a condition on the sense.

Landau (2017) argues from the other flank that the literature has been captured by perfectionism, with standards so demanding that ordinary lives fail them by construction, and that meaning under imperfection is the real article. The account agrees and supplies the mechanism: purpose is constituted wherever the modality engages real differences (§§2–3); the engagement needs neither cosmic endorsement nor a threshold of achievement (§4.1). The perfectionist standard is one more orthotic overlay mistaken for the constitution of the thing it grades. Benatar (2017) presses the opposite flank, a pessimism on which the ledger of any life weighs negative whatever purposes it contains. The dispute is first-order, conducted within the field by the comparative machinery of Chapter 15, so the constitution account neither settles it nor stands or falls with its outcome.

6.4 Nagel and the absurd

Nagel's (1971) celebrated argument deserves its own treatment rather than being subsumed under the nihilism analysis, because Nagel's absurdist is not nihilistic. She feels the force of her commitments vividly; Nagel explicitly says that the recognition of absurdity should not lead us to abandon our projects. What she has done is detect a specific evaluative difference: the difference between the significance her commitments have from the first-person perspective and the insignificance they have from the external standpoint. This is itself a detection of evaluative difference — registered by the evaluative

modality. Nagel's absurdity is not the absence of purpose but the detection of a tension between two evaluative perspectives.

This explains why Nagel is right that absurdity should not lead to despair. The absurdist's modality is fully functional, and what she has detected is not the absence of the sense but a meta-evaluative tension available to a self-conscious system that can adopt both perspectives and register the conflict between them. The tension is real, and it complicates purpose without eliminating it, since object and sense remain. What has been added — a second-order awareness that the external standpoint does not endorse what the internal one constitutes — is itself an evaluative engagement, and therefore itself a form of purpose.

Where the present account parts company with Nagel is in the inference from "no ultimate justification" to "arbitrary". Purpose does not require ultimate justification; it requires the evaluative modality's oriented engagement with non-evaluative differences, a sense directed at an object. The absence of cosmic endorsement therefore does not make our commitments arbitrary; it makes them response-dependently constituted rather than cosmically authorised. The evaluative differences are real, the commitments based on them are grounded in the structural architecture of conscious evaluation, and the absence of an external guarantee does not diminish the differences the modality constitutes.

7 Objections and Replies

7.1 The triviality objection

Objection. The chapter establishes only that creatures with emotions care about things. This is trivially true.

Reply. The objection confuses the content of the conclusion with its derivation. The conclusion may indeed be obvious as a psychological observation, while the derivation connects purpose to the analysis's foundational results and yields consequences that are not obvious: the bipartite structure (§2), the two constitutive forms with orthotic purpose as derivative (§4), the three-case diagnosis of nihilism (§5), and the

division of the meaning-of-life impasse (§6). A familiar conclusion can have important consequences when its basis is made explicit.

What the derivation buys. The consequences can be named, and naming them completes the reply, because each is a downstream result this book actually derives and rival accounts of purpose do not deliver in this form.

The saturation argument (Chapter 20, §3) turns on the object/sense structure, since purpose persists only while the object supplies real differences for the modality to engage, which is why unbounded time threatens saturation — whereas a desire-satisfaction account, on which purpose is the meeting of preference and world, predicts no such cap for a being whose stable desires stay met. The telic/atelic mechanics inherit the same structure, since an atelic object is one whose differences are not consumed by engagement, which converts Setiya's observation into a mechanism. And the fit-criticism of the grass-counter (§6.3) requires the bipartite split to state, since fit is the accuracy of the sense's registration of its object, a criticism an objective-list account can only simulate by writing the verdict into the list.

A rival account may capture the familiar conclusion; what it does not generate are these three results, which are where the derivation delivers. Williams stated the saturation, Setiya the telic/atelic contrast, Wolf the demand for fit; what the architecture adds is the mechanism each was missing, and a chapter whose yield is a mechanism for results others stated claims that yield and no more.

7.2 The mortality objection

Objection. If purpose is constituted by the evaluative modality's engagement, and the modality ceases at death, then all purpose is temporary, so Craig's (2008) argument that mortality renders life absurd remains unanswered.

Reply. Purpose requires the modality's engagement, and permanence is no part of the requirement. The object may be permanent or temporary, while the sense operates in the present: a sunset is beautiful while it is experienced, and the beauty is not diminished by its

ending. A friendship is meaningful while it endures, and its meaning is not retroactively erased by the death of a friend.

Duration is therefore a feature of the object, and the sense imposes no such condition. Craig's argument assumes that significance requires permanence of the object, whereas the analysis holds it requires engagement of the sense — a claim this reply only separates from Craig's; the argument against the permanence premise, on the modality's own evidence, is made at Chapter 20 (§2.2); the reply here is confined to keeping duration and significance apart. A further point is developed in Chapter 20: immortality may itself exhaust purpose through evaluative saturation, the progressive diminishment of the sense as every possible object is repeated without limit.

7.3 The purpose-without-God objection

Objection. The present account may show that purpose is possible without God, but it does not show that purpose is equally rich without God (Cottingham, 2003).

Reply. The chapter is partially sympathetic. If religious practice engages the modality — through communal empathic experience, contemplative practices that heighten evaluative sensitivity, rituals that orient the participant toward differences she might otherwise overlook — it may indeed provide resources for purpose. But the resources work because they engage the sense rather than because a deity authorises the engagement, and secular practices of comparable intensity — deep friendships, artistic creation, scientific inquiry, care for the suffering — provide the same structural conditions for both object and sense.

7.4 The response-dependence objection

Objection. Purpose is constituted by responses that are themselves response-dependent. This seems to undermine the "objective" element the chapter claims to preserve.

Reply. The "objective" element is not the sense (which is indeed response-dependent) but the object: the non-evaluative differences the modality responds to. The non-evaluative differences a purpose engages

(the tissue damage, the deprivation, the physical facts of a situation) are real and mind-independent; the suffering is not among them, being a felt state and therefore on the response side (Chapter 14, §6.4). What is response-dependent is the sense — the evaluative significance of those differences. This is the secondary-quality model: wavelengths are mind-independent; colours are response-dependent; the conjunction is neither purely objective nor purely subjective. Purpose has the same structure: the object is mind-independent; the sense is response-dependent; the conjunction constitutes purpose. The experience machine (Nozick, 1974) is the sense with a counterfeit object; the fit criterion of §6.3 says why its purposes fail: full registration of what the object contains dissolves them, with no appeal to a stance-independent order of worth.

8 Conclusion

Purpose is not something the universe gives us; it is not something we manufacture from nothing. It is something that happens when a conscious system with an evaluative modality encounters a world of differences and responds. The response, the sense, is involuntary; the world, the object, is not of our making. But the conjunction is real; it is ours.

The chapter yields three results, all constructed.

The bipartite structure of purpose. Purpose has an object (the target) and a sense (the emotional orientation toward it), paralleling the bipartite architecture of aesthetic experience, empathic experience, and certainty — the structural signature of the evaluative framework applied to purpose.

Aesthetic and empathic purpose. Purpose is generated within the evaluative hierarchy and takes two constitutive forms, aesthetic and empathic including self-empathic, with orthotic purpose derivative: the standard provides the object while the aesthetic or empathic response provides the sense — an instance of Chapter 16's grip argument (§3.1). The hierarchy that generates purpose is involuntarily constituted, so purpose is real without being self-authored in the ultimate sense: a product of bounded authorship.

The meaning-of-life impasse divided. The subjective, objective, and hybrid positions are partial descriptions of a single bipartite phenomenon, the sense and the object of purpose. Wolf's formula is confirmed and structurally explained; Metz's cluster conditions are unified; Frankfurt's caring is located within the bounded-authorship architecture; de Beauvoir's situated freedom anticipates the structural account; Nagel's absurdity is reinterpreted as a meta-evaluative tension that complicates purpose without eliminating it. Nihilism admits of three structurally distinct cases — evaluative silence, negative engagement, and the purposive act of self-destruction — and is evaluative diminishment rather than philosophical discovery.

If purpose is constituted within bounded authorship, and does not require an external source to be real, what happens when it meets the theological claims that it requires God's existence, cosmic endorsement, or eternal duration? The next chapter applies the present account to these questions, along with the failure of divine retribution's desert basis and the case for restorative over retributive justice.

Grades and Dependencies

1. Stakes. All three results are *constructed* (Chs. 14, 15, 18). Kalanithi and Frankl are cited for the shape of the mechanism, not as evidence of it (§3.2). The fit criterion concedes its residue: a grass-counter who survives full registration keeps his purpose, an outcome Wolf's objectivism forbids (§6.3). The diminishment diagnosis is scoped to experiential nihilism; the error theorist is met on the merits (Ch. 14), not diagnosed (§5.4). The Premise Register carries the chain.

References

Benatar, D. (2017). *The human predicament: A candid guide to life's biggest questions*. Oxford University Press.

Camus, A. (1942/1955). *The myth of Sisyphus* (J. O'Brien, Trans.). Vintage.

Cottingham, J. (2003). *On the meaning of life*. Routledge.

Craig, W. L. (2008). The absurdity of life without God. In *Reasonable faith* (3rd ed., pp. 65–90). Crossway.

de Beauvoir, S. (1947/1976). *The ethics of ambiguity* (B. Frechtman, Trans.). Citadel Press.
Frankfurt, H. G. (2004). *The reasons of love*. Princeton University Press.
Frankl, V. E. (1946/2006). *Man's search for meaning* (I. Lasch, Trans.). Beacon Press.
Kalanithi, P. (2016). *When breath becomes air*. Random House.
Kauppinen, A. (2012). Meaningfulness and time. *Philosophy and Phenomenological Research*, 84(2), 345–377.
Landau, I. (2017). *Finding meaning in an imperfect world*. Oxford University Press.
Metz, T. (2013). *Meaning in life: An analytic study*. Oxford University Press.
Nagel, T. (1971). The absurd. *Journal of Philosophy*, 68(20), 716–727.
Nozick, R. (1974). *Anarchy, state, and utopia*. Basic Books.
Ratcliffe, M. (2015). *Experiences of depression: A study in phenomenology*. Oxford University Press.
Scheffler, S. (2013). *Death and the afterlife*. Oxford University Press.
Setiya, K. (2017). *Midlife: A philosophical guide*. Princeton University Press.
Smuts, A. (2013). The good cause account of the meaning of life. *Southern Journal of Philosophy*, 51(4), 536–562.
Taylor, R. (1970). *Good and evil*. Macmillan.
Wielenberg, E. J. (2005). *Value and virtue in a godless universe*. Cambridge University Press.
Wolf, S. (2010). *Meaning in life and why it matters*. Princeton University Press.

CHAPTER 20

Purpose and Punishment

Presupposes: the dissolution of creation ex nihilo (Vol. I, Ch. 9); the response-constitutive account of value and the commitment result (Ch. 14); the primary/added value taxonomy and field-enrichment (Ch. 15); the great moral conflation (Ch. 16); the unified causal model and perspectival asymmetry (Ch. 18); the evaluative constitution of purpose (Ch. 19).

Establishes: that purpose requires neither God nor immortality; the failure of divine retribution's desert basis, by two independent routes; restorative over retributive justice; autonomy-preserving universalism.[1]

In brief. Can anyone be justly punished for what they did not ultimately author? Purpose and punishment are two faces of one coin: the modality that constitutes meaning without cosmic endorsement is the one whose involuntary delivery makes ultimate accountability impossible. Against Craig: significance requires neither permanence, endorsement, nor duration — a sunset's beauty is not diminished by its ending, and immortality invites evaluative saturation — Makropulos's tedium, given its mechanism. Against divine retribution: the wrong of eternal damnation is that it is punishment — a response to a desert no agent possesses and a value no register supplies — not that it is eternal. A judge who understood the architecture would not punish; he would educate. The terminus — every soul reconciled — is proposed for the companion God alone, as an aim rather than a guarantee.

---

1 Introduction

Can anyone be justly punished for what they did not ultimately author? This chapter applies the results of the preceding chapters to two theological claims that have shaped the lives of billions: that purpose requires God, and that God is justified in punishing those who choose wrongly. The results it applies are the evaluative constitution of purpose, with its bipartite structure of object and sense, and bounded authorship as the structural condition of any possible agent.

Purpose and punishment belong together, and the connection is structural rather than merely thematic.

The same evaluative modality that constitutes purpose (Chapter 19) is the modality whose involuntary delivery constitutes the bounded authorship that undermines absolute moral responsibility (Chapter 18). They are two faces of one coin: the first shows what the evaluative modality makes possible, meaning without cosmic endorsement, while the second shows what it makes impossible, ultimate accountability for choices the agent did not ultimately author.

The chapter argues four things. Purpose does not require God's existence, cosmic endorsement, or immortality (§§2–3). The theological claim of religious dissatisfaction rests on a misunderstanding of infinity (§4). Divine retribution lacks the desert basis it requires, given the bounded authorship of any possible agent; the strongest theological defence, Swinburne's self-separation, fails on both its premises; Plantinga's free will defence presupposes a degree of freedom the structural architecture does not provide; and religious obedience admits of four structurally distinct forms, only one of which constitutes moral goodness in its own right (§5). Finally, the structural analysis supports restorative over retributive justice and, in the divine case, education over punishment altogether (§6).

2 Purpose Without God

2.1 The structural argument

Purpose is constituted by the conjunction of an object, the target toward which the agent is oriented, and a sense, the evaluative modality's emotional engagement with that target, as Chapter 19 established. None of the structural conditions mentions a deity. Purpose obtains wherever these conditions are jointly satisfied — wherever a conscious system with a functioning evaluative modality identifies an object and constitutes an oriented response toward it — regardless of whether the universe was designed, whether the differences are cosmically endorsed, or whether a creator intended the system to have purpose.

The first section of this book established that the grounding terminus is unintelligent and that creation ex nihilo is structurally incoherent (Chapters 4, 6, 9, 10). God cannot be the origin of existence from non-being, and whether or not existence depends on a God (Chapter 24, §4.1), purpose arises from the structural architecture of conscious evaluation, not from the intentions of a cosmic designer.

2.2 Craig and the absurdity of life without God

Craig's (2008) argument is the most developed contemporary version of the theistic claim and deserves sustained engagement. Craig identifies three conditions he claims are jointly necessary for life to be meaningful: lasting consequences, cosmic endorsement, and personal survival. A godless universe satisfies none. The argument is serious because it gives voice to a real evaluative orientation — a longing for permanence, endorsement, and survival — that many people share, whether or not they share Craig's theological commitments.

The first strand is the absence of lasting consequences. Craig argues that in a godless universe all human achievements will eventually be erased, by personal death, by extinction, by the heat death of the cosmos, so nothing we do has lasting significance.

The response identifies the assumption: that significance requires permanence. The assumption is itself an evaluative claim the modality does not support, since most of what matters to most people — a child's first steps, a conversation with a friend, a problem solved — matters in the present, without reference to cosmic duration, and a sunset's beauty is not diminished by its ending (Chapter 19, §7.2). Craig conflates two claims: that significance is real, and that it requires permanent objects. The first is true; the second fails.

Scheffler (2013) develops a striking thought experiment: imagining that humanity will go extinct thirty days after one's death. Scheffler argues that this prospect would drain present activities of significance far more than the prospect of one's own death, suggesting that what matters is not personal immortality but the continuation of the human project. The present account explains Scheffler's finding: the evaluative

modality constitutes significance through its engagement with non-evaluative differences, and many of those differences presuppose the continuation of evaluative beings who can register them. The "afterlife" that matters for purpose is not personal survival but the persistence of the evaluative community: the ongoing existence of beings whose evaluative modalities can register the differences in virtue of which one's engagement is significant.

The second strand is the absence of cosmic endorsement. Craig argues that without a God who values what we value, our valuations are subjective preferences projected onto an indifferent universe. The response draws on the response-constitutive account (Chapter 14), since evaluative differences are response-dependent realities on the secondary-quality model, no more unreal than colours. Craig's demand for cosmic endorsement is a demand for a further layer of derived value — not merely that the modality constitutes significance, but that a cosmic being confirms the constitution — so the demand adds a condition purpose does not require. A mother comforting a frightened child constitutes real significance through her empathic engagement, whether or not God witnesses the act.

The third strand: the inevitability of personal annihilation. Section 3 addresses this strand.

2.3 Swinburne and the value of a calling

Swinburne (2004) offers a more nuanced version: God's existence provides human life with a point it would otherwise lack, the opportunity to develop one's character in response to a divine calling. The response operates at two levels.

First, the "calling" is a form of purposive orientation, an evaluative target the modality is oriented toward. But the calling's purpose is constituted by the modality's oriented response to it rather than by its cosmic origin. A person who feels called to alleviate suffering is oriented toward a real evaluative target, the difference between suffering and its absence, and the orientation constitutes purpose

whether the calling comes from God or from the person's own developmental history.

Second, Swinburne's argument depends on what the value taxonomy (Chapter 15) calls the confusion between total value and primary value. Swinburne is right that the total value of a calling may be greater if God exists, since the derived value of divine authorship is a real addition. But the total value's being greater does not mean the primary value is absent without it.

Metz (2019) provides the most comprehensive recent analysis of the relationship between God and meaning, distinguishing several ways in which God could contribute to meaningfulness: as creator, as moral legislator, as object of a relationship, as guarantor of cosmic justice. Metz argues that the relational contribution is the most defensible. The structural account agrees that a relationship with a being of extraordinary evaluative richness would contribute real meaning, which is what the companion God of Chapter 24 provides. What it denies is that this relational contribution requires God to be the creator or the moral legislator.

3 Immortality and Evaluative Saturation

3.1 The duration assumption

A common religious argument ties purpose to the afterlife: if we all die and there is no heaven, what is the point of existence? The argument assumes that purpose requires infinite duration, and the present account dissolves the assumption. Purpose requires the evaluative modality's oriented engagement with real differences rather than the infinite continuation of that engagement. The modality operates in the present, constituting the oriented difference between better and worse, desirable and undesirable, now, and the constitution is purpose regardless of how long it continues. Duration determines how long purpose obtains.

A forced choice makes the point vivid: annihilation on one side, an eternity of complete solitude on the other. The comparison shows where worth is located. Whether a life spans a day, a thousand years, or

forever, its worth is carried by the engaged responses that fill it, and duration without them adds nothing.

3.2 Williams and evaluative saturation

Williams (1973) develops the counter-intuitive consequence of the duration assumption in his discussion of Elina Makropulos, who is granted immortality at the age of forty-two. Three hundred years later she has become intolerably bored, emotionally flat, and indifferent to experience. Williams argues that her condition is a feature of immortality rather than a contingent failure of imagination.

The value taxonomy (Chapter 15) provides the mechanism. Primary value diminishes through habituation, since the evaluative modality's oriented response to a stimulus weakens as the stimulus is repeated, because the contrast between the stimulus and the subject's prior experiential field diminishes. Added value exhausts its associations, since over infinite duration every possible derived and consequential association would eventually be formed, explored, and exhausted. Total value, the composite of primary and added, therefore approaches zero as duration approaches infinity. This is evaluative saturation: the erosion of the contrast on which evaluative engagement depends.

One premise of this argument is an empirical wager. That the supply of real novelty is bounded is a claim about the universe, since whether it affords finitely or infinitely many distinguishable states is a question of physics, entangled with the temporal disjunction of Chapter 5, and the chapter does not settle it. The saturation argument accordingly holds on the finite branch and is offered conditionally there, while §7.2 states what survives on the other branch.

Fischer (1994) provides the most sustained philosophical response, arguing that Williams's argument depends on the assumption of a fixed character, since an immortal person with sufficient flexibility could continually develop new interests and thereby avoid exhaustion. Fischer's point is stronger than a fixed-character reply allows, since memory is finite too: an immortal who forgets re-encounters, and a bounded field that clears makes room for the same difference to register

again, so bounded novelty drives total value to zero only if the memory that would recognise the repetition is unbounded, which a bounded substrate forbids. The saturation argument therefore needs either an unbounded memory, which it cannot have, or the weaker claim that survives without one: within the span a memory holds, the field fills and its novelty is spent, and across spans the immortal repeats — total value cycling, with the tedium Williams describes being the registration of the cycle as a cycle. The chapter takes the weaker claim and marks it. In a finite universe with a finite number of distinguishable states, the supply of real novelty is bounded, so infinite time will exhaust it within every span a memory can hold, which is what the weaker claim needs and all it claims.

Cottingham (2003) argues that engagement with an infinitely rich God could provide inexhaustible evaluative novelty. The answer: this is possible in principle. A being of infinite diversity could sustain eternal evaluative engagement. But this defence trades on infinite diversity, the same infinity that is incompatible with the divine simplicity classical theism attributes to God. The defence works only for a God of infinite diversity that classical theism rejects. The irony is sharp for the classical tradition, and only for it, since the open theist and the personalist do not hold the simplicity doctrine it turns on: the theological claim that immortality with God provides inexhaustible purpose is coherent only on a conception of God that the dominant traditions deny.

The physics behind the wager. The physics behind the finite-novelty premise should be named, because the premise is an empirical wager and the wager has identifiable stakes. The bound is of the Bekenstein kind: a physical system confined within finite energy and radius carries at most a finite information content (Bekenstein, 1981), so a post-mortem career realised in any bounded physical substrate has finitely many distinguishable states to cycle through. On the cosmological side, analyses of long-run information processing in an ever-expanding universe approaching a de Sitter horizon find that only finitely many thoughts can be completed (Krauss & Starkman, 2000), against Dyson's (1979) earlier argument that an open universe might permit unending, ever-slowing experience; Dyson (2001) replied that

the objection binds only digital processing, a reply not conceded, so the wager is live. On eternal-inflation or cyclical scenarios the supply of novelty may be unbounded at the global scale, and the premise then falls back on its conditional form: what saturates is a bounded experiencer's field, whatever the universe at large does. The saturation argument is therefore strongest under de Sitter futures and bounded substrates, and the theological work it does in §3 inherits that grade.

3.3 Finitude and evaluative intensity

Finitude may be a structural condition for certain forms of evaluative richness. The knowledge that time is limited makes present engagement more precious: scarcity intensifies the evaluative modality's response. A conversation that could be the last is experienced with greater evaluative intensity than one among infinitely many. This is essentially Heidegger's (1927/1962) insight about Being-toward-death: the awareness of mortality is not a defect of human existence but an existential structure that intensifies present engagement. The analysis provides the mechanism: finitude intensifies evaluative response because scarcity is itself a difference that the modality registers. Mortality is not merely compatible with purpose; it is, for finite beings, one of the conditions under which the most intense forms of purpose are constituted.

Becker (1973) argues that the denial of death is the primary engine of human culture, since mortality anxiety generates "immortality projects" designed to transcend it. The present account accommodates this, because mortality anxiety is the evaluative modality's negative response to the prospect of evaluative cessation, a real and powerful oriented difference. But the response itself is a form of purposive engagement rather than its negation, since the person who builds an immortality project is constituting purpose through her engagement with the project. What the present account adds to Becker is the observation that the anxiety is mislocated: the evaluative modality responds to the prospect of evaluative cessation as though permanence were a condition for present significance, and the analysis shows it is not.

People of faith may struggle to see it, though belief in God is neither unique nor necessary for purpose. God can serve as a dominant emotional anchor and a powerful source of meaning, while purpose can be found wherever emotional resonance arises: in family, friendship, art, science, or a walk beneath the stars. The account does not adjudicate between hierarchies, since a life oriented toward God and a life oriented toward artistic creation are both purposive, provided the modality is engaged with real non-evaluative differences.

Wielenberg (2005) argues that value and virtue are possible in a godless universe because the conditions for moral goodness do not depend on a divine source. The analysis agrees and provides what Wielenberg's argument lacks: an explanation of why the conditions are sufficient. Evaluative differences are response-constituted realities rather than brute facts, located in the relational structure between the modality and its input rather than in the input alone.

Cottingham's (2003) more moderate position is that theistic commitment provides resources for meaning — communal worship, contemplation, a narrative framework — that secular alternatives may not fully replicate. The account is partially sympathetic, on the ground Chapter 19 (§7.3) states: the resources are real, and they work because they engage the modality rather than because a deity authorises the engagement.

4 Religious Dissatisfaction

4.1 Religious dissatisfaction and the misunderstanding of infinity

"Without God there can be no true happiness." Beliefs of this form teach their holders to expect dissatisfaction, since joy is referred to the afterlife and the present life is read as incomplete by design. The expectation rests on a misunderstanding of infinity entangled with the doctrine of God's infinite nature, and the reasoning runs: "God is infinite, and I need God; therefore I must need something infinite."

But the two infinities (Chapter 8) are structurally opposite conditions. Infinite diversity is the endless extension of internal structure; infinite equality is the dissolution of all structure. If the God in question is infinitely diverse, the felt lack the belief reports is a desire for inexhaustible evaluative engagement — real, legitimate, and satisfiable wherever the modality meets a field rich enough to sustain it, with a deity as one candidate among others. If the God in question is infinitely equal, the lack admits of no divine filling at all, because zero complexity has no evaluative content to engage.

The remedy is therefore not the acquisition of faith but the cultivation of the engagement that faith, at its best, was always meant to provide: deeper empathic connection, richer aesthetic experience, more purposive orientation toward the real non-evaluative differences that populate one's life. Whatever sits at the top of a person's evaluative hierarchy — God, family, justice, creative work, or self-respect — becomes her dominant purpose (Chapter 19, §3.2). God can serve as a powerful dominant purpose, while the structural conditions for purpose are satisfied by any engaged hierarchy rather than only a theistic one.

If purpose does not require God, what about justice? The Abrahamic traditions hold that God is the ultimate moral judge: that the righteous are rewarded and the wicked punished, if not in this life then in the next. This section applies the unified causal model and the perspectival asymmetry (Chapter 18) to this doctrine.

Where the argument stands: purpose has been shown to require neither God nor cosmic endorsement nor immortality (§§2–3), and the felt lack that religious dissatisfaction reports has been traced to a confusion between the two infinities (§4). What follows turns to punishment, and asks what desert a divine judge could be answering to.

5 Divine Retribution

5.1 The structural argument

Chapter 18 established three results. The evaluative field is populated by subconscious processes consciousness does not control,

for any conscious system, physical or non-physical. Intentional behaviour shares the chain architecture of unintentional behaviour, differentiated only by the evaluative modality. And freedom is phenomenologically real from the first-person present while metaphysically determined from the third-person retrospective perspective.

Divine retribution presupposes absolute moral responsibility (Chapter 18, §7.2): complete control over the factors that determined the choice, over the evaluative hierarchy that populated the deliberative process rather than only over the process itself. That is a presupposition no secular philosopher holds, so what the chapter delivers against the live theistic defences is the luck objection in Pereboom's (2005) form together with the value route of §6.

The value taxonomy (Chapter 15) sharpens the diagnosis. The choice is determined by total value, the composite of primary and added, and the agent chose neither component, since the primary responses were constituted involuntarily by the modality and the added-value profile is a product of developmental history and involuntary association. The total value that determines her choice is therefore the product of an evaluative architecture she did not design.

Adams (1999) argues that "horrendous evils", sufferings so severe that they threaten to render the sufferer's life meaningless, present a challenge to any theodicy grounded in proportionate response to free choices. Adams's analysis converges with the bounded-authorship thesis, because if some evils are so severe that they overwhelm the agent's capacity for evaluative engagement, then the choices the agent makes under their influence are products of an evaluative hierarchy distorted by conditions she did not create.

Consider two people raised in radically different environments: one in a loving religious household where God is associated with warmth, security, and community, the other in an abusive religious household where God is associated with fear, punishment, and control. Both "freely choose" to accept or reject God, while the evaluative architectures from which they choose are radically different, and neither

person authored her architecture. Holding both agents to the same standard of eternal accountability ignores the structural conditions under which the choices were made.

5.2 Conditional love

Freedom under sanction. It is often said that God has left us free to accept or reject Him. But a freedom whose every exercise is met with judgement is freedom only in name, since where one option is answered with eternal sanction the agent is not so much free to choose as free to incur the penalty.

The deeper objection, however, is not that the stakes render the freedom illusory but that the sanction has no warrant to begin with. Retributive punishment, the imposition of suffering because it is deserved, presupposes that the agent is the ultimate author of the self from which the wrongful choice issued, and §5.1 and Chapter 18 deny that any agent is. The wrong of eternal damnation is therefore not, in the first instance, that it is eternal; it is that it is punishment. The absence of the authorship that desert requires withdraws the warrant for retribution at every magnitude, so the infinite form merely extends without limit a sanction that was unfounded at any.

The same analysis reaches the claim that divine love is unconditional. A regard that is withdrawn on the outcome of a judgement the agent lacked the ultimate authorship to meet is conditional in the strict sense, since it is extended on a condition the agent was never positioned to control. To make eternal punishment the consequence of falling short is to make acceptance contingent on a performance that an involuntarily constituted hierarchy may or may not have allowed, which is regard conditioned on an outcome the agent did not author rather than unconditional regard. If hell is on the table, the love that tolerates it is conditional, where the condition is one no agent meets, or fails to meet, by anything that traces to him.

The point does not reach every version, and does not need to. On the self-exclusion models (§6; Kvanvig, 1993; Walls, 1992) nothing is withdrawn, since the love continues to be offered and the damned

decline it, so the regard is unconditional in the sense the objection requires; escapism (Buckareff & Plug, 2005) sharpens the model by keeping the door open without end, so that the offer of reconciliation is never withdrawn. What those models cannot decline is the prior argument. A soul that turns away turns away from within a hierarchy it did not author, and the suffering consequent on the turning is not thereby deserved — which is why the wrong of damnation is that it is punishment rather than that it is imposed.

An omniscient being who understood the structural architecture of consciousness, seeing the unified chain from the third-person retrospective perspective with perfect clarity, would recognise that the desert retribution presupposes — an ultimate authorship of the self — belongs to no conscious agent. The infinite stakes of divine retribution, eternal punishment for wrong choices and eternal reward for right ones, are on this view unfounded rather than merely excessive: not a proportionate response overshot, but a response to a desert that is not there.

The strongest theological defence reconstructs eternal punishment not as externally imposed torture but as the natural consequence of free rejection of God. Swinburne (1998) argues that hell is the condition of a soul that has definitively turned away from the source of all goodness. God does not send anyone to hell; the agent sends herself. The punishment is the natural consequence of the rejection, just as loneliness is the natural consequence of pushing away everyone who loves you.

The self-separation defence. Swinburne's defence is the most sophisticated available, and the analysis must engage it at its foundation. The argument depends on two premises: that the agent's rejection of God is free in the full sense, and that the consequences of this free rejection are proportionate.

The unified causal model engages the first premise directly. The choice to reject or accept God is itself a product of the agent's evaluative hierarchy: her emotional responses to the religious evidence she has encountered, her developmental history with religious authority, her subconsciously constituted orientation toward or away from the

divine. The two-environment illustration (§5.1) makes the point vivid, since a person raised in a loving religious household and a person raised in an abusive one "freely choose" from radically different evaluative architectures. Both choices reflect real evaluative hierarchies, and both are real at the phenomenological level, while neither person authored the hierarchy from which the choice flows.

The perspectival asymmetry (Chapter 18) identifies the mechanism. The rejection is phenomenologically free, since the agent really is evaluating, and metaphysically determined, since it is produced by a chain of evaluative forces she did not initiate. Swinburne's defence operates at the first-person level, where the rejection is free, and draws consequences at the metaphysical level, where eternal separation is said to be proportionate, so the two levels deliver divergent reports.

Eternal consequences for such a choice therefore answer to a desert the choice cannot carry, as a sentence premised on the offender's authorship of his own will is undercut once that authorship is shown to be absent. Duress is the familiar, local case in which conditions outside the agent's making defeat the desert a sanction would require; bounded authorship is the general case, obtaining for every agent and every choice.

The universality objection. The generalisation invites a stock compatibilist reply; it should be met where the analogy is used. A mitigating condition that obtains for every agent and every choice, the reply runs, mitigates nothing — a currency issued to everyone is worth nothing — so bounded authorship, being universal, cannot do the exculpatory work that duress, being exceptional, does.

The reply mistakes the logic of the two notions it compares. Mitigation is comparative and admits of a curve; desert is not. Duress defeats desert in the local case not because it is rare but because it severs the premise a retributive sentence requires, the agent's authorship of the will from which the act issued, and a severing that turns out to be universal does not restore the premise; it universalises its failure. What is graded on a curve is practical responsibility, which the analysis preserves intact and which continues to distinguish the coerced from the

willing, the reflective from the impulsive (Chapter 18, §7.1). Basic desert is pass–fail, and bounded authorship fails every agent alike, which is not the inflation of an excuse but the discovery that the sentence's premise was never available to be satisfied.

The second premise, proportionality, is undermined by the same analysis. Even if the agent's choice is "hers" in the practical sense that it reflects her evaluative hierarchy, the hierarchy's constitution is bounded, and the ethereal mind argument (Chapter 18) shows this holds for any possible agent, physical or non-physical. Eternal consequences for a choice produced by bounded authorship therefore answer to a basic desert that bounded authorship withholds, regardless of how sincere or reflective the choice appeared from the inside.

The problem of evil in its traditional form, the tension between divine omnipotence, omniscience, omnibenevolence, and the existence of suffering, does not arise if there is no omniscient, omnipotent, omnibenevolent creator at the origin of existence. But Plantinga's (1974) free will defence deserves engagement on its own terms, because the analysis provides a critique that goes beyond the standard objections.

Plantinga argues that a world containing creatures who freely choose good is more valuable than a world of automata, and that God cannot create free creatures while guaranteeing they always choose well. The suffering in the world is therefore the cost of a freedom God could not eliminate without destroying the value it produces.

The analysis targets Plantinga's central premise — not that free will exists, since it agrees that choice is phenomenologically real, but that the creatures' freedom is of the kind that justifies the cost. If the agent's evaluative hierarchy was involuntarily constituted by processes she did not control, then the disobedience Plantinga invokes is the product of an evaluative architecture the agent did not design rather than of the absolute free choice the defence requires.

The placement matters, because the objection belongs to a lineage the defence has met before. Pereboom (2005) argues that rejecting libertarian free will removes the free-will theodicy, since creatures who

lack that freedom cannot carry the cost the defence asks them to bear, and the present analysis reaches his conclusion by its own route.

Where it differs from the standard incompatibilist reply, which denies free will under determinism, and from the logical problem of evil, which finds a contradiction among God's attributes and suffering, is in the ground it occupies. It accepts the reality of choice — the agent chose, and the choice reflects her hierarchy — while denying that this reality suffices for the absolute responsibility the proportionality claim requires, because the hierarchy from which the choice flows was not itself chosen. This is the hard-incompatibilist point of Chapter 18 brought to bear on the defence, with the substrate-generality of the ethereal mind argument closing the libertarian escape routes a determinism-based version leaves open. The "risk" Plantinga identifies, that God cannot guarantee free creatures will choose well, is therefore not a risk freely accepted by the creatures; it is a risk imposed on creatures whose evaluative hierarchies were constituted by processes they did not control.

Van Inwagen (2006) has developed the "expanded free will defence", on which the first human beings, in turning from God, lost the protection they had enjoyed from the hazards of the natural order, so that what human beings suffer from natural causes traces to a misuse of free will as moral evil does; Plantinga (1974) had allowed that natural evil might be the work of fallen angels. The analysis applies to both expansions as it applies to Plantinga's original, since the evaluative hierarchies of the first humans and of the angels were themselves involuntarily constituted, and each "rebellion" flowed from a hierarchy its bearer did not author.

Stump (2010) offers a theodicy that shifts the ground from freedom to love: suffering is the occasion for redemptive love rather than the cost of freedom. The engagement is more sympathetic than with Plantinga, because Stump grounds the justification in empathic concern. But two problems remain. Stump presupposes a creator God who designed the world in which suffering occurs, and whether a person responds to suffering with deepened compassion or embittered withdrawal depends on her evaluative hierarchy, which was not authored by the sufferer.

The insight that suffering can, under the right conditions, deepen empathic engagement is preserved; the theological claim that this deepening was divinely orchestrated is not.

Four forms of obedience. If morality is constitutively empathic (Chapter 16), what is the moral status of religious obedience, compliance with divine commands out of reverence for God? The three-goodness taxonomy (Chapter 14) provides the diagnostic.

Empathic-alignment obedience occurs when the agent obeys because she empathically grasps the command's content: she understands why cruelty causes suffering, and her evaluative modality responds to that suffering. The command and the empathic response converge. This is the best form of obedience, since it combines orthotic compliance with empathic engagement.

Fear-based obedience occurs when the agent obeys because she fears divine punishment. The preference for cruelty is overridden by fear rather than transformed. This is shallow persuasion (Chapter 21): the underlying preference survives, suppressed by a stronger emotional response.

Reward-based obedience occurs when the agent obeys because she seeks divine reward: paradise, blessing, salvation. The evaluative engagement is with the reward, not with the persons the command affects. The structure is identical to fear-based obedience: the preference is overridden, not transformed.

A fourth form must be added, because it is the one the tradition itself puts first. Love-of-commander obedience occurs when the agent obeys from love of God, neither to escape punishment nor to secure reward, but because the beloved asks. The tradition is right that this differs in kind from fear and reward, since it is empathic in structure — a real oriented response, engaging and transformative where fear and reward merely suppress — and nothing in the account denies its reality or its beauty.

What settles its moral standing is its object. The response is directed at the commander's state — His will, His pleasure, the relationship —

rather than at the persons the commanded conduct affects, while morality's subject matter is conduct's bearing on those it affects (Chapter 16, §2.1).

Love-of-commander obedience therefore inherits its moral credit from the alignment of the command's content: where God commands kindness, the lover of God does good at one remove. Where command and cruelty coincide, love of the commander is the devoted perpetrator's structure — the compliant perpetrator of Chapter 16 (§5.2) with love in place of procedure, the most sympathetic member of that family and still within it. The nobility of the motive is real; it is not the moral fact.

Only empathic-alignment obedience constitutes moral goodness in its own right, since love-of-commander obedience is moral where, and because, the commands it follows are empathically aligned. A theology that evaluates persons by their obedience — whether the obedience is empathically aligned, love-based, fear-based, or reward-based — is conflating orthotic compliance with moral goodness. This is the great moral conflation (Chapter 16) in its most intimate religious form.

6 Restorative Justice and the Empathic Framework

Two independent routes. Retributive justice holds that wrongdoing deserves punishment proportionate to the wrong, so the punishment is backward-looking. The analysis faces retribution with more than a limit on its degree, meeting it on two grounds that do not lean on each other.

The first concerns the value the suffering is said to have. Basic desert — the non-instrumental desert that would make suffering owed for its own sake — claims that the wrongdoer's suffering is good in itself, and the metaethics has a test for claims of that shape: ask who it is good for. Developed below, the test returns the vacancy result for the response-independent reading and the anchor's verdict for the response-dependent one, and it uses nothing the applied sequence does not already rest on (Chapter 14).

The second concerns the responsibility the suffering is said to answer to. Basic desert requires the ultimate authorship of the self that Chapter 18 denies to any agent, on the sourcehood premise Chapter 18 (§8.1) names as contested. Either ground removes retribution's basis, so a reader who refuses the sourcehood premise meets the first ground unchanged. What is withdrawn is therefore not a quantum of proportionate punishment, leaving a smaller measure intact, but the desert basis as such. Practical responsibility is preserved, and with it the forward-looking response examined below; backward-looking retribution is not.

Restorative justice holds that the appropriate response is the repair of the relationship between the wrongdoer, the victim, and the community rather than punishment. The response is forward-looking, since it aims to transform the wrongdoer's evaluative hierarchy so that the preference for the wrongful conduct is itself changed — deep persuasion (Chapter 21) applied to the specific conflict the wrongdoing created.

Duff (2001) develops a communicative theory of punishment, on which punishment is justified as a form of moral communication: the community's expression to the offender that her conduct was wrong, inviting recognition and repentance. Tadros (2011) argues that punishment is justified by the duty a wrongdoer incurs to protect her victim, and others, from further harm. Both accounts are structurally aligned with the empathic framework, since Duff's communicative punishment is deep persuasion, engaging the offender's evaluative modality through moral communication, while Tadros's victim-centred approach prioritises the empathic relationship between offender and victim over the orthotic satisfaction of a retributive standard.

A third family should be named as surviving untouched: quality-of-will accounts, on which blame responds to the meaning of the agent's conduct, the ill will or indifference it expresses, without invoking basic desert at all (Scanlon, 2008; McKenna, 2012). The value route's test must be run on this family too, or the instrument cuts unevenly: blame is unpleasant to its object, so a blamee's registration returns "bad" as a sufferer's does. What distinguishes blame that survives from suffering

that does not is that blame is address — a communication whose point is the blamee's uptake, and which succeeds as blame only if it is understood — whereas retributive suffering is imposed for its own sake and succeeds whether or not anything is understood. The first is a registration offered to a register; the second is a registration inflicted on one. The test "good for whom?" returns, for blame, the practice of address and both its parties, since address is something done with the blamee and not to her, and for retribution, the persecutor alone; the distinction is drawn against the instrument, and the family survives on it. Communicative, victim-centred, and quality-of-will accounts neither need the desert eternal retribution requires nor lose anything to its removal. The lineage runs to P. F. Strawson (1962), whose reactive attitudes ground the practice, with Darwall (2006) supplying its second-personal grammar; both survive on the same terms.

One further family sits at the fork the two routes create, and its placement completes the taxonomy. The fair-play tradition (H. Morris, 1968) grounds punishment in a distribution: the law-abiding carry the burden of self-restraint, the offender takes the benefits of the system while shedding the burden, and punishment restores the balance.

Read as a fairness rationale, the account uses no basic desert, since the debt it collects is anchored in the registrations of those who bore the restraint, and the response it licenses is answerable to the same mutually informed field as every other claim about burdens (Chapter 15, §4.4). It therefore survives beside the communicative, victim-centred, and quality-of-will accounts, at whatever weight the field assigns it.

Read as a desert rationale — the offender's suffering owed because taking the advantage made suffering his due — it falls to both routes above, since the owing is the vacant clause under a distributive description and the taking issued from a hierarchy the taker did not author. Berman (2011) draws the line within retributivism that this fork tracks, separating consequentialist versions, on which deserved suffering is a good the state has reason to bring about, from non-consequentialist versions, on which the desert claim does its work without routing through the goodness of outcomes, and Husak (1992) argues from inside the retributive literature that desert alone cannot

carry the justification of punishment — a conclusion the present analysis reaches with the further claim that nothing else restores what desert was to supply.

Braithwaite (1989) provides the most influential criminological evidence: reintegrative shaming, which engages the wrongdoer's empathic modality rather than merely inflicting pain, produces lower recidivism than retributive punishment. The structural prediction aligns with Braithwaite's empirical finding, since restorative approaches that engage the offender's evaluative modality through empathic confrontation with the victim's suffering produce more durable behavioural change than punitive approaches that merely override the offender's preferences through fear.

The structural analysis supports restorative over retributive justice on three grounds. It is aligned with deep persuasion, aiming to transform the preference rather than override it. It is compatible with bounded authorship, presupposing no absolute control over the hierarchy. And it engages the modality in its empathic deployment, bringing the wrongdoer into contact with the victim's suffering so that the hierarchy is transformed from within.

The strongest retributivism. This leaves the strongest form of retributivism, which rests on neither deterrence nor reform: the claim that the guilty simply deserve to suffer, that their suffering is good in itself and owed independently of any further end (Kant, 1797/1996; M. S. Moore, 1997). The forward-looking case for restoration does not touch this position, which grants that punishment serves no purpose and embraces it regardless.

It is met instead at two points, each sufficient on its own. First, basic desert of this non-instrumental kind requires the ultimate authorship of the self that the bounded-authorship thesis withholds (Chapter 18), so there is no desert for retribution to answer to. Second, "deserved suffering is good in itself" faces the fork on which the metaethics is built. Read response-independently, good in a way that would obtain apart from every response, the claim is the vacancy

result's target by name (Chapter 14, §3.6): a clause marking no contrast between worlds.

Read response-dependently, the claim does not vanish but changes hands. The retributivist's satisfaction at the wrongdoer's suffering is a real oriented response, and the account does not deny it; it weighs it. A response that takes another's suffering as its object with positive orientation is the persecutor's structure (Chapter 16, §6.3), and what it constitutes is what every such response constitutes: that the suffering is good to the one approving.

What it cannot constitute is that the suffering is owed, because the moral fact about the sufferer is anchored in the sufferer's own registration (Chapter 14, §6.4), and hers returns the opposite verdict. The intrinsic-desert retributivist must therefore locate the good of punishment either in a value the vacancy result has shown to be unconstituted, or in an orientation the anchor has already outranked — the persecutor's, offered as a theory of justice.

The response-dependent reading also admits of a test the account has used before. The informed-registration anchor (Chapter 14, §6.5) asks whether a registration survives accurate acquaintance with its object, and the retributivist's satisfaction is a registration whose object is the wrongdoer's suffering. Does it survive the victim's full registration of that suffering?

The restorative-justice trials have put the question directly. Across the controlled comparisons reviewed by Sherman and Strang (2007), victims who met their offenders face to face reported less desire for violent revenge, and more satisfaction with justice, than victims processed conventionally. The satisfaction at suffering thinned under acquaintance with the sufferer, which is what the account predicts of a registration whose object it had not fully registered. Where a victim's satisfaction does survive acquaintance, the account does not dismiss it; it weighs it as one registration against the sufferer's, in the mutually informed field of Chapter 15, and what such fields settle on is the chapter's evidence for restoration over retribution: two screens and no third, rather than a scale that reads "deserved".

Nor can he retreat from constitution to commitment, to the claim that desert-talk at least means the response-independent owing, since the commitment result (Chapter 14, §A.3) forecloses that retreat for "owed" as for "wrong": a vacant clause conditions nothing, so the discourse of desert is answerable only to its contrastful remainder, which the anchor has already adjudicated.

The case against retribution accordingly stands on two routes, and the relation between them should be stated.

	The value route	**The authorship route**
Rests on	The response-constitutive account of value: the vacancy result (§3.6), the anchor (§6.4), the commitment result (§A.3) — nothing the applied sequence does not already use (Chapter 14)	Bounded authorship and its sourcehood premise (Chapter 18)
What it removes	The value the suffering is said to have: a suffering good for no register is, in the model of Chapter 14 (§A.5), a field with no body	The responsibility condition: no agent is the ultimate source of the self its desert requires
If the other route fails	Were ultimate authorship available, the responsibility condition would return and the value claim would stand where it stood	Were the response-independence clause contentful after all, this route would still deny the source
Whom it meets	The compatibilist about responsibility, who refuses the sourcehood premise	The retributivist who grants bounded authorship — met by both routes

Each route removes the desert basis by itself. What survives the surrender of one premise is the result by the other route, and only the surrender of both would give retribution its basis back.

The nearest development in the literature reaches the verdict by the second route alone. Caruso (2021) argues from scepticism about basic desert to the rejection of retributive punishment, with a public-health model supplying the forward-looking remainder. The convergence is recorded together with the difference: the value route carries the result here independently, so the verdict survives a reader who keeps ultimate sourcehood, while the surviving practices are those the chapter names — communicative, victim-centred, quality-of-will, restorative — rather than a quarantine model alone.

The theological placement. Traditions that propose finite, remedial punishment — purgatory in Catholic theology, Gehenna as purification in some Jewish traditions, temporary punishment in some Islamic traditions — stand closer to restorative justice than traditions that propose eternal retribution, as do the self-exclusion models on which the damned are not punished but withdraw from God of their own accord (Kvanvig, 1993; Walls, 1992; cf. §5.2).

But the bounded-authorship thesis does not challenge merely the infinite escalation of retributive consequences, as though finite retribution were sound and only its unbounded form excessive. It challenges retributive desert as such. The target of the argument is the eternal retributive hell of the strong tradition; annihilationist and remedial models, which abandon eternal retribution, already fall outside it; and the model toward which the argument's own logic points is neither, but universalism — the eventual reconciliation of every soul.

The universalism the argument points to has a theological lineage of its own, and the convergence should be recorded together with the divergence of route. Within the Christian tradition the position runs from Origen's apokatastasis to its contemporary defenders. Talbott (2014) argues that a God who loves each creature cannot will the final loss of any, so eternal damnation contradicts the tradition's own first premise. D. B. Hart (2019) argues that an eternal hell is morally

unintelligible on any candid accounting of finite guilt, and that a creation purchased at the price of one soul's endless torment could not be called good.

The present argument reaches the shared destination from outside the tradition's premises, needing neither the perfection of divine love nor the moral grammar of creation, only bounded authorship and the educative means developed below — a convergence of differently premised routes of the kind this book treats as evidence of structure. And the theological universalists retain the creator God whose coherence Chapter 9 dissolves, so their conclusion survives here with its ground exchanged.

Why the divine case differs. For the divine case in particular, even the forward-looking justification that survives for human institutions falls away. Human justice retains restorative and protective measures because human agents cannot reach into a wrongdoer and remake the hierarchy from which the wrong issued; they must work through sanction, constraint, and the slow labour of moral communication. An omnipotent God is under no such limitation. What restoration aims at — the transformation of the evaluative hierarchy so that the preference for the wrong is itself changed — such a God could accomplish directly, and without suffering.

The decisive qualification is that this is education, with no part for coercion. To overwrite a soul from without, replacing its responses wholesale, would destroy the very autonomy that makes it a person; that is brainwashing. It is what a God who valued persons would not do.

Autonomy-preserving transformation is instead deep persuasion at its limit: a working through the soul's own evaluative responses, so the change is one the soul undergoes and comes to own rather than one imposed upon it. Bounded authorship is what makes this coherent against the free-will defence, since there is no libertarian self-authorship for such transformation to violate, and divine education is then one more formative input to a hierarchy that was never self-originated, differing from ordinary moral formation only in its perfect

responsiveness. A God so placed has no remaining reason to impose suffering at all, so punishment, finite or infinite, is gratuitous.

The constructive consequence is conditional. If souls survive death and are not punished, what remains is for them to be brought to peace, and the only autonomy-respecting route to that peace is education: the gradual turning of every soul toward reconciliation. That reconciliation is not evaluative quiescence but engagement with a being of inexhaustible diversity (Chapter 24), the one form of unending life §3 found immune to evaluative saturation. This is the universalism the argument points to — not salvation imposed, but the eventual reconciliation of all. A divine judge who understood the architecture would educate rather than punish.

Containing the consequence. One consequence of the argument must be contained before anything is banked. If an omnipotent God could bring every soul to reconciliation by autonomy-preserving education, the question why the education waits — why any pre-mortem suffering exists at all — arises at once, since the same grant that makes punishment gratuitous makes the world's suffering unaccounted for. Nothing in this section is proposed on that grant.

The chapter's dialectical position should therefore be stated with care. Omnipotence is granted arguendo, to the theist who asserts it, and the omnipotent educator is a reductio addressed to him: if there is an omnipotent God, punishment is without warrant, because a better means is by hypothesis available. The reductio proposes nothing; it removes the retributive theist's footing on his own assumptions, and the problem of evil the grant reopens stays with the assumption's owner, since the chapter has taken nothing from it.

On the framework's own theology the grant is taken up by one survivor and declined by the other. The sustainer of Chapter 24 (§4.1) holds the causal order in being and has the means the conditional assumes, so its force falls on him undiminished. The companion God has no such means, since he cannot remake hierarchies or the natural order at will, and he answers for no suffering he lacks the power to prevent; his educative relation is persuasion across time rather than

transformation on demand (Chapter 24, §6.5). The universalism this section proposes is accordingly the companion's alone. It is conditional on the survival of persons beyond death, and, since the companion cannot remake a hierarchy on demand, it is a hope with a mechanism rather than a prediction: restoration as what such a God keeps working toward across time, never an outcome he guarantees (Chapter 24, §§5.2, 6.5).

7 Objections and Replies

7.1 The cosmic significance objection

Objection. Local purpose is not cosmic significance. The theist's claim is not that people without God cannot care about things but that their caring has no ultimate grounding: that in a godless universe, even the deepest engagement is a brief flicker in an indifferent cosmos.

Reply. The desire for cosmic significance is real and should not be dismissed. But significance is constituted by the modality's engagement with real differences rather than by a further layer of cosmic endorsement, so demanding endorsement from the cosmos adds a condition the account shows unnecessary. The desire itself is the modality's oriented character reaching beyond its structural conditions, where the reaching is itself evaluative engagement, and therefore a form of purpose.

7.2 The Williams counterargument

Objection. An immortal being who develops radically new interests need not experience evaluative saturation.

Reply. The question is whether real novelty is inexhaustible, which is §3.2's question and carries §3.2's answer: on the finite branch the supply is bounded and infinite time exhausts it. Fischer's response is strongest in a universe of infinite diversity and weakest in a finite universe approaching heat death.

The reply, carrying the hedge of §3.2, is accordingly conditional. On the finite branch saturation wins and the flexible immortal runs out

of world. On the infinite branch Fischer wins the duration question, and even there what survives for the chapter is the conditional Craig must deny: purpose requires engagement rather than permanence, so mortality threatens nothing that infinite duration would rescue.

7.3 The proportionality defence

Objection. Even if freedom is bounded, proportionate punishment is still possible. Finite punishment for finitely free choices may be perfectly just.

Reply. The objection conflates retributive desert with forward-looking sanction. What bounded authorship withdraws is the desert basis of retribution as such (§6); finite retribution is no better grounded than infinite, only smaller. What is coherent is the forward-looking, restorative response that does not rest on desert, and in the divine case not even that (§6). Practical responsibility is preserved throughout; retributive desert, at any magnitude, is not.

7.4 The mercy objection

Objection. Theological traditions incorporate divine mercy alongside divine justice, accommodating bounded authorship through doctrines of grace.

Reply. Traditions that qualify retribution with mercy have already moved in the direction the analysis recommends, since they have recognised, implicitly, that absolute responsibility is too strong a standard. The analysis provides a principled foundation for the theology of mercy: if absolute responsibility is structurally unavailable, then mercy is not a divine concession to human weakness but the only just response to the conditions under which human beings actually make choices.

7.5 The moral motivation objection

Objection. Without divine punishment, what motivates moral behaviour? Fear of God may not be the noblest motivation, but it is effective.

Reply. The objection confirms the chapter's analysis. Fear of divine punishment is shallow persuasion: it overrides the preference for wrongdoing without transforming it. Deep persuasion (the transformation of the preference through empathic engagement with the suffering wrongdoing causes) is more stable and more morally adequate. A society that relies on divine fear for moral behaviour is a society in managed conflict, not one in structural peace — the condition in which the conflict has been resolved rather than managed (Chapter 21).

7.6 The practical value of religion objection

Objection. Religious traditions provide practical resources for meaning that secular alternatives struggle to match. Dismantling the theological justification may leave people without the infrastructure that sustains purpose.

Reply. The chapter agrees that religious traditions provide valuable practical resources and does not argue for their dismantlement. The practical resources work because they engage the evaluative modality, so they can in principle be preserved even if the theological metaphysics is revised. What the chapter denies is that the resources derive their value from a divine source rather than from their capacity to engage the evaluative modality with real differences.

7.7 The religious experience objection

Objection. A feeling of divine forgiveness is evidence of divine mercy.

Reply. A feeling is evidence only of a feeling: a vivid experience of divine forgiveness is evidence that the agent experienced being forgiven (a real and potentially transformative evaluative event) not that God exists, forgives, or grounds the forgiveness. The transformation is real; the theological interpretation is a further claim the feeling itself does not establish.

8 Conclusion

The chapter yields three results, all constructed.

Purpose without God or immortality. Craig's three strands each fail because they add conditions the analysis shows are not required: permanence, endorsement, and infinite duration. Swinburne's "calling" is a purposive orientation whose primary value is independent of its cosmic origin. Williams's argument converges with the account, since evaluative saturation is the mechanism that explains the tedium of immortality — and the irony stands: sustaining purpose through immortality would require a God of infinite diversity, the very infinity incompatible with classical theism's simple God.

The failure of divine retribution's desert basis. Divine retribution presupposes absolute moral responsibility, a form of freedom Chapter 18 shows unavailable to any possible agent, and a value in the sufferer's suffering that no register supplies (Chapter 14); either lack removes its basis.

The Swinburne self-separation defence fails on both premises, since the rejection is phenomenologically free while metaphysically determined, while the eternal consequences answer to a desert no agent possesses — the defence failing not because the penalty is too large but because none is owed. Plantinga's free will defence presupposes a degree of freedom the architecture does not provide, which is the hard-incompatibilist point his existing responses do not address. Religious obedience admits of four structurally distinct forms, and only empathic-alignment obedience constitutes moral goodness in its own right, love-of-commander obedience borrowing its standing from the alignment of what is commanded.

Restorative justice and universal reconciliation. The structural analysis supports restorative over retributive justice, and in the divine case dispenses with punishment altogether. What is abandoned is not a measure of retribution but its desert basis, since even an omnipotent agent who could transform a hierarchy by autonomy-preserving education would have no warrant to punish at all — the wrong of

eternal damnation being that it is punishment rather than merely that it is unending.

The chapter's constructive terminus is the educative reconciliation of every soul, universalism of the autonomy-preserving kind, proposed for the companion God alone: conditional on survival, and offered as an aim that God pursues by persuasion across time rather than an outcome he can guarantee (§6). This stands on two independent routes, the value route through Chapter 14 and the authorship route through Chapter 18, each of which removes the desert basis by itself, so only the surrender of both would give retribution its footing back.

If purpose does not require God, and divine retribution has lost its desert basis, what happens when these insights meet the conditions for democratic peace? The next chapter develops the three modes of conflict resolution — enforcement, shallow persuasion, and deep persuasion — and argues that only deep persuasion produces structural peace.

Grades and Dependencies

1. Conditions. The first three results are *constructed*; universalism is *proposed* for the companion God alone: conditional on post-mortem survival, and an aim pursued by persuasion across time rather than a guaranteed outcome (Chapter 24, §§5.2, 6.5); the omnipotent educator of §6 is a reductio addressed to the retributive theist and proposes nothing. The retribution case stands on two independent routes — Chapter 14 (vacancy, anchor, commitment) and Chapter 18 (bounded authorship) — each sufficient; the value route carries the result for a reader who refuses the sourcehood premise, and the victim-satisfaction finding (§6) is an empirical stake. The saturation argument is an empirical wager (§3.2). The Premise Register carries the chain.

2. What survives the desert argument (§6). Removed: basic desert, hence retribution at any magnitude. Untouched: practical responsibility; communicative (Duff), victim-centred (Tadros), quality-of-will (Scanlon, McKenna); fair-play on its fairness reading (H. Morris), its desert reading falling with the routes. Outside the target: annihilationist and remedial models. Met at the prior step: self-exclusion models — the suffering still

undeserved. Converging by another route: Talbott and Hart, ground exchanged.

References

Adams, M. M. (1999). *Horrendous evils and the goodness of God*. Cornell University Press.

Becker, E. (1973). *The denial of death*. Free Press.

Bekenstein, J. D. (1981). Universal upper bound on the entropy-to-energy ratio for bounded systems. *Physical Review D*, 23(2), 287–298.

Berman, M. N. (2011). Two kinds of retributivism. In R. A. Duff & S. P. Green (Eds.), *Philosophical foundations of criminal law* (pp. 433–457). Oxford University Press.

Braithwaite, J. (1989). *Crime, shame and reintegration*. Cambridge University Press.

Buckareff, A. A., & Plug, A. (2005). Escaping hell: Divine motivation and the problem of hell. *Religious Studies*, 41(1), 39–54.

Caruso, G. D. (2021). *Rejecting retributivism: Free will, punishment, and criminal justice*. Cambridge University Press.

Cottingham, J. (2003). *On the meaning of life*. Routledge.

Craig, W. L. (2008). The absurdity of life without God. In *Reasonable faith* (3rd ed., pp. 65–90). Crossway.

Darwall, S. (2006). *The second-person standpoint: Morality, respect, and accountability*. Harvard University Press.

Duff, R. A. (2001). *Punishment, communication, and community*. Oxford University Press.

Dyson, F. J. (1979). Time without end: Physics and biology in an open universe. *Reviews of Modern Physics*, 51(3), 447–460.

Dyson, F. J. (2001, March 13). Is life analog or digital? *Edge*. https://www.edge.org/conversation/freeman_dyson-is-life-analog-or-digital

Fischer, J. M. (1994). Why immortality is not so bad. *International Journal of Philosophical Studies*, 2(2), 257–270.

Hart, D. B. (2019). *That all shall be saved: Heaven, hell, and universal salvation*. Yale University Press.

Heidegger, M. (1927/1962). *Being and time* (J. Macquarrie & E. Robinson, Trans.). Harper & Row.

Husak, D. (1992). Why punish the deserving? *Noûs*, 26(4), 447–464.

Kant, I. (1797/1996). *The metaphysics of morals* (M. Gregor, Trans.). Cambridge University Press.

Krauss, L. M., & Starkman, G. D. (2000). Life, the universe, and nothing: Life and death in an ever-expanding universe. *The Astrophysical Journal*, 531(1), 22–30.

Kvanvig, J. L. (1993). *The problem of hell*. Oxford University Press.
McKenna, M. (2012). *Conversation and responsibility*. Oxford University Press.
Metz, T. (2019). *God, soul and the meaning of life*. Cambridge University Press.
Moore, M. S. (1997). *Placing blame: A theory of the criminal law*. Clarendon Press.
Morris, H. (1968). Persons and punishment. *The Monist*, 52(4), 475–501.
Pereboom, D. (2005). Free will, evil, and divine providence. In A. Dole & A. Chignell (Eds.), *God and the ethics of belief: New essays in philosophy of religion* (pp. 77–98). Cambridge University Press.
Plantinga, A. (1974). *God, freedom, and evil*. Eerdmans.
Scanlon, T. M. (2008). *Moral dimensions: Permissibility, meaning, blame*. Belknap Press of Harvard University Press.
Scheffler, S. (2013). *Death and the afterlife*. Oxford University Press.
Sherman, L. W., & Strang, H. (2007). *Restorative justice: The evidence*. The Smith Institute.
Strawson, P. F. (1962). Freedom and resentment. *Proceedings of the British Academy*, *48*, 187–211.
Stump, E. (2010). *Wandering in darkness: Narrative and the problem of suffering*. Oxford University Press.
Swinburne, R. (1998). *Providence and the problem of evil*. Clarendon Press.
Swinburne, R. (2004). *The existence of God* (2nd ed.). Clarendon Press.
Tadros, V. (2011). *The ends of harm: The moral foundations of criminal law*. Oxford University Press.
Talbott, T. (2014). *The inescapable love of God* (2nd ed.). Cascade Books.
van Inwagen, P. (2006). *The problem of evil*. Oxford University Press.
Walls, J. L. (1992). *Hell: The logic of damnation*. University of Notre Dame Press.
Wielenberg, E. J. (2005). *Value and virtue in a godless universe*. Cambridge University Press.
Williams, B. (1973). The Makropulos case: Reflections on the tedium of immortality. In *Problems of the self* (pp. 82–100). Cambridge University Press.

CHAPTER 21

Conflict and Deep Persuasion

Presupposes: the primary/added value taxonomy (Ch. 15); the great moral conflation (Ch. 16).

Establishes: the three modes of conflict resolution; their hierarchical relationship; deep persuasion as the only source of structural peace.

In brief. Most of what passes for peace is not peace at all — it is managed conflict, held in place by force, incentive, or fatigue, and stable only while its overrider holds. A conflicting preference can be addressed at exactly three points: the capacity to act on it (enforcement), the calculation around it (shallow persuasion), or the preference itself (deep persuasion) — and there is no fourth. Only the third leaves nothing suppressed, nothing waiting on a change of circumstance to re-express itself; the strict definition of peace is forced, not stipulated. Peace and freedom turn out to be one limit approached from two directions. Deep persuasion is field-enrichment between agents, and its empirical signature is durability: a shift that survives the removal of what sustained it. The clock of peace advances one transformed preference at a time.

---

1 Introduction

Most of what passes for peace is not peace at all. It is managed conflict: the suppression of disagreement through force, incentive, or fatigue. A society in which dissenters are silenced is subdued. A society in which dissenters comply because the cost of defiance exceeds the cost of submission is stable. Neither is peaceful. Structural peace requires something neither force nor incentive can produce — the transformation of the preferences that generated the conflict in the first place.

The chapter commands nothing. Before the structural question can be asked, a prior one must be met: does the chapter command that peace be sought? It does not. The analysis is descriptive rather than prescriptive, and its central claims are conditional.

If aligned social peace is the goal, then deep persuasion is the only route to it, and the conditions developed in the next chapter are what that route requires.

Whether to adopt the goal is left open, and deliberately so, because the goal is itself a target of the kind this chapter analyses: to come to want social peace is to undergo deep persuasion rather than to accept an axiom. The empathic account of Chapter 16 does not compel the aim; it explains why the persuasion so reliably succeeds, since an agent whose empathic modality engages the suffering that conflict produces is, on exposure and reflection, moved toward its relief. The chapter therefore earns its organising value the way the next will argue all values must be earned: by persuasion rather than by authority. What follows maps the mechanism for those the persuasion has reached, and commands nothing of the rest.

What, then, is the peace so reached? Strictly, it is the limit at which emotional alignment is maximal: the state in which each person can act on her preferences because no one's preferences stand against them, so doing what one wants and doing what others would have one do coincide. The freedom this names is unconditioned by others, since nothing is being suppressed; it is not licence, because the wants that would collide with others have been dissolved in the very alignment that produces it — the empathic constitution of morality (Chapter 16) having already shaped each agent's wants to include the others'. Difference is not erased but freed to act without collision, so the suffering conflict produces falls away with the opposition that generates it.

Real communities only approach this limit, so the degree of alignment a society reaches is the degree of freedom its members enjoy. The limit is a limit, and the argument needs less than it. What §2.1 states, and what the results of this chapter and the next require, is alignment at some level: agreement on the framework under which the lower-level dispute is conducted. The complete alignment the limit names is the direction in which the clock of peace points (§4) rather than a condition a society must meet for the analysis to apply.

Peace and freedom are one limit approached from two directions, within the domain the identity concerns, which is constraint by other agents. Nature, scarcity, and the limits of one's own body constrain action too; they are not conflict, and no alignment dissolves them. The identity is between peace and social freedom, not between peace and omnipotence. Within that domain the only constraint on action is another's opposed preference, so alignment removes the opposition and the completion of peace and of social freedom coincide. This reframes a tradeoff much political theory takes for granted, on which social harmony is purchased at the cost of individual liberty: here the harmony and the liberty arrive together, each the measure of the other.

The republican objection. One tradition contests the identity rather than the tradeoff; it should be named where the claim is made. On the republican account (Pettit, 1997; engaged at Chapter 22, §2), freedom is non-domination rather than the absence of opposed preference: the contented slave of a kindly master holds no opposed preference and is unfree, since the master retains the power to interfere at will.

The reply is not stipulation but Chapter 15 (§6.6). Her contentment is an adaptive preference sustained by foreclosed alternatives, and the informed-registration anchor (Chapter 14, §6.5) returns the opposed preference the domination suppressed, so the case is not alignment but its counterfeit. Where alignment is real, surviving her accurate and complete registration of what is done to her, the retained power is a capacity she does not want exercised. The republican will not accept that description of his complaint, since for him domination is unfreedom now; what the reply achieves is location — the accounts divide on whether a standing power to interfere, unexercised and unwanted, is itself a constraint on action or a prediction about one, and the identity of peace with freedom holds on the second reading only.

This freedom-through-alignment also places peace within the book's metaphysics, and it forestalls a misreading the framework's own vocabulary invites. Given the partition of Chapter 3 it is tempting to picture peace as an equalisation of the evaluative field, the dissolution of the differences in which conflict consists. The picture is inverted, and Chapter 6 says why: equality is the unintelligent terminus, the state of

zero distinction that can host no mind and so no value, so a peace that equalised the field would abolish the parties rather than reconcile them.

Conflict is not difference but opposed difference: two evaluative orientations bearing on one behaviour from contrary directions, the one preferring its occurrence and the other its omission. Resolution removes the opposition rather than the difference. When two agents come to prefer the same of the same behaviour their preferences align without fusing, since there remain two evaluative processes, one in each person, reaching a shared direction rather than collapsing into one.

What peace preserves is therefore not a difference of direction, because on the contested behaviour the directions coincide, but the irreducible difference of bearer: two minds that remain two even in perfect agreement. Peace never approaches the equality pole, because the plurality of persons is never crossed; it approaches only the alignment of distinct processes. This is the content of a difference without opposition — the social form of the book's commitment to difference, not its dissolution.

The question this chapter asks is not what political institutions best promote peace, which is a question for political science. The question is structural: what must be true of a process of conflict resolution for the resolution to be stable? The answer is that conflict resolution admits of exactly three modes, that only one achieves lasting stability, and that the three modes stand in a hierarchical relationship grounded in the evaluative priorities of the agents involved.

The argument defines conflict and peace structurally (§2), then develops a three-mode taxonomy of conflict resolution — enforcement, shallow persuasion, and deep persuasion — and argues for its exhaustiveness (§3). It establishes the hierarchical ordering among the three modes and the progressive trajectory of the clock of peace (§4), examines lawmaking and law enforcement as distinct operations, including the vulnerability of law to emotional fluctuation (§5), and engages Habermas's ideal speech situation and the deliberative democracy tradition of Fishkin and Dryzek (§6).

2 The Structure of Conflict and Peace

Conflict is, strictly, the co-presence of two preferences bearing on a single behaviour from contrary directions: one agent prefers that the behaviour occur, another that it not occur. The preferences are moral preferences — evaluative orientations constituted by the empathic modality, generating dispositions that move agents toward or away from specific conduct.

Two features of this structure govern what follows. The negative preference is a preference in full rather than the absence of one, since to prefer that an action be omitted is to occupy a determinate position, whether the omission is secured by acting otherwise or by doing nothing at all. Declining to act, where another prefers the action, is therefore the holding of one pole rather than a neutral standing-aside.

And because a behaviour either occurs or does not, two agents engaged with the same behaviour hold preferences that are exhaustively concordant or contrary — aligned on a single pole or set against each other — with no intermediate relation available. Preferences that merely fail to collide are not an intermediate case, since they bear on different behaviours, and where the behaviours differ no shared object exists over which a conflict could arise. The binary is not a simplification imposed for tractability but the contrastive individuation of Chapter 2 at the level of preference: a preference for an action is the preference it is only against the preference for its omission, as a content is the content it is only against what it excludes.

Scarcity is not conflict. One clarification forestalls a natural objection to the definition's scope. Not every situation in which two agents cannot both have what they want is a conflict in the present sense. Two agents who both prefer that a single indivisible good come to them — one kidney, one appointment, one seat — are, at the level of behavioural preference, aligned: each prefers that the needy be helped, and neither prefers that the other be denied. Their predicament is scarcity rather than opposed preference, since the frustration is imposed by the world's not affording both satisfactions rather than by either holding the pole the other rejects.

The distinction matters because the two call for different treatment, and the chapter's central claim is confined to the second. Scarcity is addressed by allocation, by procedures for dividing or sequencing a good the parties do not evaluatively oppose, and a just allocation can leave both parties in full evaluative alignment, wanting for each other the very thing the shortage denies. Opposed preference is addressed by the three modes below, because there the parties want incompatible things of the same behaviour, and no division of a good dissolves the opposition.

The two can of course combine, since scarcity often generates opposed preference, as when the parties come to prefer their own claim over the other's and the shortage hardens into rivalry. When it does, the opposed-preference component is what the analysis treats, and deep persuasion — here, the parties' coming to weigh each other's need as their own — is what resolves it. The residual allocation problem remains, while ceasing to be a conflict in the strict sense: it is a distribution to be arranged among agents who no longer stand against one another. The peace–freedom identity of §1 is accordingly a claim about opposed preference rather than about scarcity, since alignment removes the opposition that constrains action without conjuring a second kidney. What it secures is that where goods must still be divided, they are divided among agents whose preferences no longer collide.

Peace is the presence of shared behavioural preferences, or the absence of conflict. The definition is strict, with the strictness derived rather than stipulated. A community is peaceful when no one dislikes anyone else's behaviour, when everyone already behaves as everyone else wishes they would. This is the positive condition set out in §1: with alignment maximal, each can act on her preferences without standing against another's, so peace is at once the absence of conflict and the fullest exercise of freedom.

Galtung's (1969) distinction between negative peace, the absence of direct violence, and positive peace, the absence of structural violence and the presence of social justice, maps onto the present account directly. Negative peace corresponds to enforcement, since the capacity for violent action has been removed; positive peace corresponds to deep

persuasion, since the preferences that generated the conflict have been transformed. The strict definition used here corresponds to Galtung's positive peace and identifies the structural conditions that make it possible.

Why strictness is forced. That the strict definition is the right one, and not merely one stipulation among several, can be shown rather than asserted. Consider the looser definition a critic is likely to prefer: peace as stable coexistence, in which compliance is maintained whether or not the underlying preferences have been reconciled. Galtung's negative peace and the Rawlsian modus vivendi are of this kind.

Any such state is held in place by something — by the capacity to enforce, by an incentive structure, or by fear or fatigue — and each of these is contingent. Enforcement capacity erodes, incentives shift, fear fades. When the sustaining factor weakens, the conflicting preference, which was never transformed and so never went anywhere, re-expresses itself as conflict. A coexistence sustained this way is therefore not a distinct terminal condition but a metastable one: conflict in suspension, stable only while its contingent overrider holds. By the analysis of §3 it is managed conflict, and it belongs on the conflict side of the ledger rather than the peace side.

The only condition not held in place by a contingent overrider is the one in which there is no suppressed preference left to re-express: the condition in which the conflicting preference has itself been transformed into a shared one. This is the strict definition; it is now seen to be forced rather than chosen. It names the unique fixed point of conflict resolution — the only state from which a change of circumstance produces no reversion to conflict, because nothing is being held down.

The apparent counterexample confirms this rather than threatening it. A constitutional or federal arrangement that both parties accept as fair, though neither has changed the underlying preference, is peaceful just to the extent that the parties share a preference for the arrangement itself: a shared preference at the level of the framework, reached by deep persuasion about its fairness, even where first-order disagreement

persists beneath it. To the extent that acceptance is instead merely strategic, better than the alternative given present power, it is shallow persuasion, and it reverts when the balance of power shifts.

2.1 Levels of alignment

The strict definition invites an objection with institutional weight behind it, and meeting it here fixes the definition's scope. Rawls's (1993) fact of reasonable pluralism holds that some evaluative disagreements survive ideal deliberation among reasonable agents, not through ignorance, bad faith, or failed communication, but through the burdens of judgement: differences of experience, weighting, and interpretive standpoint that no procedure eliminates. If some first-order disagreement is permanent in principle, then a peace defined as the transformation of every conflicting preference is not a distant ideal but an incoherent one, and the strict definition names a state no possible society occupies.

Alignment is levelled. The reply is that alignment is levelled, and the strict definition, correctly read, requires alignment at some level rather than at all of them. Preferences bear on behaviours, while arrangements for handling disagreement about behaviours are themselves objects of preference, and so upward. Call first-order behavioural preferences level zero, preferences over the frameworks that govern level-zero disagreement level one, and preferences over the choice among frameworks level two.

The federal case displays the mechanism. Two communities that disagree about a behaviour, and cannot be brought to agree, may nonetheless come — by deep persuasion about fairness rather than by strategic calculation — to a shared preference for the arrangement under which each decides for itself. The level-zero conflict persists; the level-one preference is shared; and the analysis of deep persuasion applies at the higher level as at the lower, since nothing is suppressed there, no overriding response holds anything down, and no change of circumstance finds a conflicting level-one preference waiting to re-express itself. A conflict is structurally resolved when there is some

level at which the parties' preferences concerning it are shared and every level below is governed by that shared preference.

The ceiling principle. The governance clause deserves its own statement, because it is what blocks the manufacture of "peace" by ascent. Agreement at some lofty level counts for nothing unless it regulates the levels where the conflicting conduct lives: the shared preference must be the one under which the lower-level dispute is conducted, constrained, and, when it flares, adjudicated. A shared abhorrence of mutual annihilation that leaves the parties' fighting just as before governs nothing and resolves nothing. A conflict is managed when every level exhibits either opposition or merely strategic acceptance.

The strict definition is thereby preserved rather than weakened, since transformed shared preference remains the unique fixed point and the complete alignment §1 named as the limit is the limit of this fixed point rather than a threshold. What is fixed is its object: the transformation peace requires is located wherever in the hierarchy it can be had, and the burdens of judgement, which concern the recalcitrance of level-zero differences, give no reason to expect failure at every level.

Rawls's own overlapping consensus, on this reading, is a special case of the account rather than a rival to it: deep persuasion succeeding at level one, about the terms of association, among parties whose level-zero hierarchies remain distinct. Rawls's insistence that the consensus be moral rather than a modus vivendi is the deep/shallow distinction of §3, drawn at the framework level.

Two consequences follow. First, the limit at which alignment is maximal is now doubly indexed: a society approaches peace both as more conflicts are resolved and as the level at which each is resolved descends — framework alignment where preference alignment is unavailable, preference alignment where it is. The clock of peace (§4) advances along both dimensions, so a society whose deepest disagreements are held within shared frameworks is at peace with respect to them in the strict sense, however lively the level-zero contest remains.

Second, the diagnosis of managed conflict gains sharpness. What condemns enforcement and shallow persuasion is not that first-order disagreement persists beneath them, since it may persist beneath peace itself, but that no level exists at which the parties' preferences are shared rather than suppressed. Agonistic democracy (§7.2) is accommodated on the same terms: the ongoing contest it prizes is level-zero difference exercised inside level-one alignment — not the elimination of politics but its housing.

3 Three Modes of Conflict Resolution

When two agents hold opposing behavioural preferences, the conflict can be resolved in exactly three ways.

The trichotomy is a partition rather than a taxonomy read off the cases, and its exhaustiveness can be derived. For a conflict to end, the colliding conduct must cease, and either the preference behind it persists or it is transformed; there is no third condition for a preference to be in.

If the preference persists, compliance is externally maintained; there are exactly two handles by which conduct can be secured against a standing preference: the agent's capacity, what she is able to do, or her calculation, what doing it costs or yields — incentive, threat, and deception included, since deception operates by corrupting the calculus. Enforcement is the capacity route; shallow persuasion is the calculation route. If the preference is transformed, the resolution is internal and self-maintaining: deep persuasion.

The exhaustiveness of the modes therefore rests on one premise a challenger must reject — that conduct contrary to a standing preference is sustained only through capacity or calculation — and the cost of rejecting it is a form of compliance that is neither compelled, nor chosen on balance, nor wanted, which is not a form of compliance at all. That only the third mode yields structural peace then follows from the strict definition of §2: a resolution is structural when no suppressed preference waits on a change of circumstance, and on the first two routes the preference, by hypothesis, persists.

The partition has neighbours in the conflict-resolution literature, recorded here with the difference marked. Kelman (1958) distinguishes compliance, identification, internalisation as processes of attitude change; Deutsch (1973) separates destructive from constructive processes of resolution; Burton (1990) grounds settlement in needs rather than negotiating positions. Each classifies observed change, whereas the three modes fall out of the persistence of a preference under the two handles a party can hold, so the exhaustiveness claimed above is structural rather than empirical.

3.1 Enforcement

Enforcement alters an agent's behaviour without changing her underlying preference. The attacker is disarmed or imprisoned. He may still want to harm you, but he is no longer able to act on that desire. Enforcement operates on capacity, not on motivation. Arendt (1970) separates power, resting on acting in concert, from violence, which can destroy power while never creating it; enforcement in the present sense operates on the violence side of that line, reaching conduct while leaving the preference untouched.

Enforcement is sometimes necessary, since an immediate threat to life cannot wait for persuasion. But enforcement alone cannot produce peace. If all disputes were resolved by enforcement, the only "peaceful" society would be one in which all dissenters have been killed: peace by extermination. Keeping dissenters alive through imprisonment or forced silence is not peace in the deeper sense, because the preference for conflict remains.

The post-1945 occupation of Germany and Japan involved massive enforcement: the military apparatus was dismantled, leaders were tried, and the capacity for aggressive action was physically removed. The enforcement was necessary, since it stopped the immediate harm; it did not by itself produce peace. What followed — reconstruction, institutional reform, cultural exchange, and decades of diplomatic engagement — was the slow work of persuasion.

The German case audited. The case can carry more weight than illustration, because it is concrete enough to run the informed-registration audit on, and running it shows the test operating. Ask of the post-war German transformation the audit's questions.

Were the preference shifts produced under concealment — of the conduct's nature, of the victims, of the record? The opposite: occupation-era policy forced exposure through documentation, trials, curricula, and the publicised evidence of the camps, so the shift was produced by registration where the prior regime had operated by its suppression.

Did the shifts revert when the sustaining pressure was removed? The occupation's enforcement apparatus wound down within a decade, though other pressures persisted — Western integration and its incentives in the Federal Republic, Soviet control in the East — so the axis returns a provisional verdict rather than a clean one. What it can record is that the preference shift deepened across the following two generations, surviving reunification and political stresses that would re-expose a merely strategic accommodation. Did the transformed preferences generalise beyond the enforced domain? They did: into constitutional culture, memorial practice, and a politics of historical responsibility no occupier demanded.

On all three axes the audit returns the deep verdict, and the contrast case sits within the same history: where exposure was curtailed and accommodation strategic, in parts of the post-Soviet space, the pattern re-expressed when the pressure lifted (Chapter 22, §7.4). The audit is thereby shown to be a usable instrument, with the taxonomy's discriminator — durability under removed sustaining conditions — an observable one.

The instrument should be stated as method and carried forward, because live cases cannot wait for the retrospect in which durability settles. Classification of a current transformation proceeds on the three axes prospectively: whether the shifts are produced under concealment or under exposure; whether they revert wherever the sustaining pressure locally relaxes; and whether they generalise beyond the enforced

domain. The axes are measurable while the case runs, the verdict they return is provisional and revisable as the record lengthens, and durability remains the criterion of which they are the running proxies.

3.2 Shallow persuasion

Shallow persuasion occurs when an agent refrains from conflicting action not because her preference has disappeared but because another emotional response temporarily dominates it. A bribe may make compliance more attractive than defiance. A threat of punishment may make the cost of acting on the preference outweigh the benefit. In each case the conflicting preference survives, overridden by a stronger emotional response.

Institutions on both sides of the line. Institutional arrangements occupy both sides of the deep/shallow line, which runs where §2.1 drew it. Constitutional frameworks, federal arrangements, negotiated compromises, and modus vivendi agreements divide into two cases.

Where acceptance is strategic — each party tolerating the arrangement because the alternative is worse for it, given present power — the arrangement is shallow persuasion at a high level of institutional sophistication: real, valuable, and reversion-prone, since the acceptance lasts while the balance that sustains it lasts. Where acceptance is real, each party sharing a preference for the arrangement as fair, the conflict has been resolved by deep persuasion at the framework level (§2.1): the level-zero disagreement persists, nothing at level one is suppressed, and no shift of circumstance finds a conflicting framework-preference waiting to re-express itself.

What the chapter notes of the strategic case is that the underlying disagreement remains and can resurface when enforcement weakens, when the incentive structure shifts, or when circumstances embolden the dissenter. Cold War deterrence (Schelling, 1960) provides a further illustration: the nuclear powers were shallowly persuaded to avoid direct confrontation by the overriding emotional response of fear. The compliance was real and lasted decades; it was management rather than peace, since the confrontation was suspended while the preferences

behind it were untouched. Cooperation sustained by reciprocity (Axelrod, 1984) sits on the same side of the line, since the strategy answers the incentive structure while the preferences behind the rivalry persist.

A second case from the same period shows the other mode. Within the Soviet sphere, national and ethnic conflicts were held down by enforcement, and when the enforcing capacity collapsed after 1991 the suppressed disagreements re-expressed themselves — which is what the taxonomy predicts of a preference that was never transformed.

3.3 Deep persuasion

Deep persuasion occurs when the conflicting preference itself is changed or removed. If you show the attacker convincing evidence that you were not involved in the wrongdoing he suspected, his motivation to harm you vanishes. The conflict is resolved at its root — not by removing his capacity to act, not by overriding his desire with a stronger one, but by transforming the desire itself. After deep persuasion, the agent no longer has the conflicting preference. There is nothing to suppress, nothing to override, and nothing to re-emerge when circumstances change.

Deep persuasion works by restructuring the agent's evaluative hierarchy, by introducing emotionally meaningful insights, perspectives, or experiences that cause the agent to re-evaluate the significance of the conduct in question. It is not merely the provision of new information, which might leave preferences unchanged, but the provision of information that engages the evaluative modality and produces a real shift in emotional response.

Stated in the vocabulary of Chapter 15, deep persuasion is field-enrichment (Chapter 15, §4.4) operating between agents rather than within one. The perspectives communicated are new secondary objects entering the other's evaluative field, each carrying added value grounded in real registrations, until the composite governing the conflicting preference tips. The mechanism is inherited rather than constructed, since what Chapter 15 derived for the deliberative deadlock

transfers to the interpersonal case, the field being enriched having the same architecture whoever supplies the enrichment. The discipline transfers with it: legitimate enrichment bottoms out in real registrations, a constraint whose work appears in §7.6.

Lederach (2005) develops the "moral imagination" in peacebuilding: the capacity to imagine oneself in a web of relationships that includes one's adversary and to envision a shared future neither party has yet experienced. On the present analysis that is the modality operating in its empathic deployment at the social scale. Bar-Tal (2013) provides the complementary social-psychological analysis: intractable conflicts are sustained by "conflict-supporting narratives" — in the present vocabulary, hierarchies in which the derived value of group identity overrides the primary empathic response to the adversary's suffering. Deep persuasion, in his terms, is then the transformation of these narratives through sustained engagement with the adversary's actual evaluative experience.

The South African Truth and Reconciliation Commission provides a concrete, if imperfect, illustration. It exposed perpetrators and victims to each other's experiences, creating conditions in which the modality could be engaged by the suffering the system had caused. There is no better illustration of the deep/shallow difference than the TRC's mixed record. Where the process succeeded it produced real shifts in evaluative response — perpetrators who came to feel the weight of what they had done, victims who came to see perpetrators as persons capable of remorse. Where it failed it produced the appearance of reconciliation: shallow persuasion dressed as deep, the forms of engagement without the substance.

3.4 The exhaustiveness of the taxonomy

The exhaustiveness derived at the head of §3 can be restated at the level of the architecture. A preference can be addressed at exactly three points: the capacity to act on it (enforcement), the relative strength of competing responses in the field (shallow persuasion), or the preference itself (deep persuasion). There is no fourth, because a preference either

persists unchanged or is transformed, and persistence admits only the capacity and calculation handles.

One putative fourth mode must be addressed directly, because it appears to resolve conflict without operating at any of the three levels. The mode is exit, in Hirschman's (1970) sense of the option a party keeps beside voice: separation, secession, avoidance, or the division of a contested domain so that each party governs its own. Voice is persuasion in the present vocabulary; exit reduces as follows. Exit seems to dissolve the conflict by removing the co-presence that the definition of conflict requires, rather than by acting on capacity, on the relative strength of competing responses, or on the preference itself.

It is not a fourth mode, since it reduces to the existing three. Where exit divides a contested object — two communities that disagree over a single rule agreeing that each locality will set its own — the original preference is not addressed at all. What is removed is the shared object over which it was opposed. This is real resolution only if the division is itself a shared preference, both parties preferring local determination to the alternatives. In that case the conflict has been resolved at the third level with respect to the arrangement: deep persuasion about the framework rather than a new mode. If the division is accepted only strategically, the first-order preference survives and re-expresses itself whenever the division is contested, which is shallow persuasion.

Where exit instead removes co-presence over an object that cannot be divided, the parties simply ceasing to interact, the opposed preferences are suspended rather than transformed, dormant until renewed contact revives them. This is management by avoidance, a limiting case of enforcement in which the capacity to act on the preference against the other is removed by removing the other from reach. On neither construal does separation resolve the conflict at a level the taxonomy does not already contain. The three modes remain exhaustive.

Where the argument stands: conflict and peace have been defined structurally, with alignment levelled and the ceiling principle fixed (§2); the three modes of resolution have been derived as an exhaustive

partition (§3). What follows orders them and shows why the ordering is one of dependency rather than of practice.

4 The Hierarchical Ordering

The three modes stand in an ordering of need, and the ordering is structural in one respect only: where survival or security is not secured, deep persuasion cannot begin. The ordering has parallels with Maslow's (1943) hierarchy of needs, since both identify a sequence in which higher-level achievements presuppose the satisfaction of lower-level conditions. The grounding differs. Maslow's hierarchy is a psychological generalisation about the order in which human beings typically pursue needs; the present hierarchy is a claim about dependency.

Deep persuasion requires the prior satisfaction of enforcement and shallow persuasion conditions, regardless of the empirical order in which individuals happen to pursue them. The claim is not that all human beings or all societies satisfy survival needs before security needs. The claim is that deep persuasion structurally requires that the conditions the other modes address, survival and security, obtain by whatever means: you cannot transform someone's evaluative hierarchy if she is dead or too afraid to engage. Where the conditions obtain untended, as between parties who were never at war, deep persuasion proceeds with neither mode having operated. Where they do not, enforcement and shallow persuasion are what secures them — which is why the modes are ordered as they are even though they need not all have run.

Enforcement addresses survival. When someone is pointing a gun at your head, the immediate evaluative priority is survival. You cannot deeply persuade someone who is actively killing you.

Shallow persuasion addresses security. Once survival is secured, the evaluative priority shifts to security — the confidence that the threat will not return tomorrow. A person cannot begin the slow work of understanding her adversary's evaluative perspective if she is still afraid he will attack her.

Deep persuasion addresses alignment. Once security is established, a deeper evaluative priority can emerge: understanding, shared values, the elimination of the underlying disagreement. Deep persuasion is the terminal mode, the one that produces actual resolution rather than management, and it depends on the prior modes having done their work.

Each instance of deep persuasion expands the territory of shared understanding, and as shared understanding expands, fewer conflicts require enforcement or shallow persuasion. The trajectory is progressive without being guaranteed. New conflicts can arise through new circumstances, new generations with differently constituted evaluative hierarchies, or new situations that trigger previously absent emotional responses. The institutional apparatus of conflict resolution must therefore be preserved as a standing capacity, not because it is currently needed but because it may be needed again. The apparatus is like a fire station in a town with no fires: the absence of fires does not justify dismantling the station.

The progression can be visualised as a clock of peace, each moment of deep persuasion turning the hand forward and expanding the territory of shared understanding. Over time, enforcement and shallow persuasion fade as they are needed less, until a full rotation is complete and peace is achieved.

The clock is a regulative ideal and makes no prediction. A society may never complete the rotation. But the direction of travel matters: a society whose clock is advancing, whose instances of deep persuasion are increasing and whose reliance on enforcement and shallow persuasion is decreasing, is moving toward structural peace even if it never arrives. The German trajectory examined in §3.1 supplies the one dated illustration the metaphor otherwise lacks (an illustration, since a regulative ideal admits no calibration) — enforcement dominant in 1945, institutional and educational persuasion carrying the load within a decade, and preference-level transformation measurable across two generations: one dated case of the hand advancing, with the post-Soviet contrast (§3.1) as the clock run in place.

5 Lawmaking and Law Enforcement

5.1 Two operations

Lawmaking and law enforcement are different operations, and their conflation produces characteristic pathologies.

Lawmaking is the process by which private moral preferences are transformed into shared public standards. Its ideal mechanism is deep persuasion: agents communicate their evaluative hierarchies, engage with the hierarchies of others, and arrive at shared preferences that are then codified as rules. The moral authority of a law derives from the quality of the persuasion that produced it. A law codifying a shared preference has deeper moral authority than a law imposed by a majority over a dissenting minority, which in turn has deeper authority than a law imposed by force.

Law enforcement operates across all three modes: it enforces through arrest, fines, and imprisonment; it shallowly persuades through deterrence; and it can contribute indirectly to deep persuasion through education. The distinction matters because lawmaking and law enforcement have different normative requirements. Lawmaking requires persuasion, because its product, shared standards, derives its moral authority from evaluative convergence. Law enforcement may legitimately use enforcement and shallow persuasion, because its function is to maintain compliance with standards already established.

Law enforcement is also a more technical act than lawmaking, since it does not require emotional involvement. An officer is not tasked with agreeing or disagreeing with legislation; she compares observed behaviour to what is written in the legal code. Even if she personally opposes a specific rule, she can still recognise whether an action violates it. The written law translates subjective moral preferences into objective behavioural standards that can be applied without evaluative engagement.

A society that uses enforcement in its lawmaking, imposing laws by force rather than establishing them through persuasion, is authoritarian. A society that uses only deep persuasion in its law enforcement, relying

entirely on transforming preferences without any coercive apparatus, cannot protect its members from imminent harm. The two operations require different mixes of the three modes, and conflating them produces these characteristic pathologies.

In practice the operation runs on mutual adjustment: each person works on the structure of the others' evaluations until a shared preference emerges, an adjustment that, done well, deepens each party's empathic grasp of the others' primary and secondary objects of evaluation (Chapter 15). A participant who comes to weigh environmental protection more heavily once she has felt the health consequences it carries for identifiable others has been persuaded in the deep sense, since her modality has responded to differences it had not previously been shown, and nothing has overridden her. Chapter 22 (§5.1) works the mechanism through in detail.

5.2 Emotion and the vulnerability of law

Because moral preferences are rooted in emotional responses, law — being built upon such preferences — is vulnerable to the fluctuations of emotion.

Consider a person who regards the death penalty as immoral and campaigns for its abolition. After years of advocacy, the government changes the law. But the day after the reform, her child is abducted and harmed by a serial offender. Overcome by grief and rage, the same person who fought to ban capital punishment now demands its reinstatement. Her evaluative hierarchy has shifted, not through new evidence or deeper empathic understanding but through the overwhelming intensity of a personal emotional response. This is not a failure of character; it is the evaluative modality operating as it always operates — constituting oriented responses whose intensity is determined by the proximity and vividness of the relevant evaluative differences.

The scenario reveals the danger of placing lawmaking in the hands of a single individual, whose emotional state may shift quickly and radically: fear or hatred toward a group one day, a law to punish them

the next. Most legal systems therefore require that several people, each with a different emotional context, agree before laws change, so laws are revised only after multiple evaluative hierarchies have been engaged, reducing the probability that a single intense response will distort the process. The more fundamental the legislation, the broader the agreement needed to amend it — as in constitutional law, where supermajority requirements and multi-stage ratification are designed to buffer the evaluative volatility the modality's emotional constitution makes possible.

H. L. A. Hart (1961) provides the jurisprudential foundation. The "rule of recognition", the master rule that identifies which other rules count as law, is itself a social practice, sustained by the convergence of officials' evaluative attitudes. On the structural account, Hart's rule of recognition is the institutional expression of a shared evaluative hierarchy among officials: a hierarchy produced by persuasion, deep or shallow, and maintained by the evaluative modality's continued oriented response to the practice.

The vulnerability is that the rule of recognition, being evaluatively constituted, can shift when the evaluative hierarchies of officials shift. The institutional safeguards against evaluative volatility — constitutional entrenchment, judicial independence, separation of powers — are therefore structurally necessary rather than merely convenient.

6 Habermas and the Ideal Speech Situation

6.1 Discourse ethics and deep persuasion

Deep persuasion overlaps with Habermas's (1990a, 1990b, 1996) ideal speech situation and discourse ethics. Both hold that consensus requires free and open communication, that a norm's legitimacy depends on the quality of the discursive process that produced it, and that compliance achieved through force differs from agreement achieved through the uncoerced exchange of reasons.

The convergence is substantial while the grounding differs. Habermas derives the conditions for legitimate discourse from the pragmatics of communication: anyone who enters an argument implicitly commits to norms of sincerity, truth-telling, and equal participation. The present framework derives them from the priority of deep persuasion, itself derived from the evaluative architecture — deep persuasion is the only mode that produces structural peace, it requires communication, and communication requires free and open speech. The conditions are therefore derived from the metaphysics of evaluative difference rather than the pragmatics of argumentation.

The difference in grounding has a consequence. Habermas's transcendental-pragmatic argument is vulnerable to the objection that one can simply refuse to argue, so that the presuppositions of argumentation bind only those who choose to engage. The present framework's derivation avoids this objection, because the conditions are derived from the requirements for peace rather than from the act of arguing. A person who declines to argue is free to do so, so her declining limits the available modes of conflict resolution to enforcement and shallow persuasion, neither of which produces peace.

The symmetry should be stated as plainly against the framework as for it: the derivation binds no one either. Its conditions are hypothetical imperatives, binding on whoever wants structural peace, and a person who declines the end escapes them as completely as Habermas's refuser escapes the pragmatics of argumentation. The advantage claimed is therefore not that the framework obligates where Habermas cannot, but that its conditional survives the refusal intact: the refuser is not caught in a performative contradiction, and needs to be caught in none, because the fact that only enforcement and shallow persuasion remain to her stands whether or not she ever argues.

6.2 Deliberative democracy: Fishkin and Dryzek

The concept of deep persuasion also converges with the deliberative democracy tradition, which holds that legitimate political decisions are those produced through inclusive, reason-giving deliberation among affected parties. Fishkin (2009) has developed deliberative polling: a

representative sample of citizens deliberates on a policy question under conditions of balanced information and moderated discussion, with the resulting opinion shifts measured.

Fishkin's empirical findings bear directly on the account. Deliberative polls consistently show significant opinion change after sustained engagement with opposing perspectives, particularly where participants are exposed to the experiences and arguments of those who disagree with them. On the structural analysis, deliberative polling is an institutional mechanism for facilitating deep persuasion at scale. The structured exposure to others' evaluative hierarchies provides the emotionally meaningful input that can produce real shifts in evaluative response.

The durability evidence. A more pointed piece of evidence bears directly on the deep/shallow distinction, because it isolates the one variable the taxonomy makes load-bearing: durability. Broockman and Kalla's (2016) field experiment found that a single canvassing conversation inviting the resident to recall and dwell on an analogous experience of her own — active perspective-taking — produced reductions in prejudice toward transgender people. The reductions persisted for at least three months and survived exposure to subsequent attack advertising.

What the intervention supplied was neither information nor incentive but the emotionally meaningful engagement of the resident's own modality with a perspective she had not previously registered. The persistence is the diagnostic feature. A shift produced by incentive or framing reverts when the sustaining factor is removed (§2); a shift produced by evaluative engagement leaves nothing to revert. That the effect outlasted a deliberate counter-message is the empirical signature the taxonomy predicts for the deep mode and denies to the shallow one. Kalla and Broockman's (2020) replications, identifying non-judgemental exchange of narratives as the operative ingredient, reinforce the reading: what transforms a preference durably is the very input §3.3 identifies as constitutive of deep persuasion.

Three qualifications keep the evidence weighed at its size. The finding's own history enforces the first: the 2016 study was undertaken as the corrective replication of an earlier canvassing result that collapsed under fraud. The subsequent experiments, while confirming durable effects, find them modest in size and uneven across issues. The magnitudes belong in the text: the 2016 study, on 501 voters, reported a reduction comparable to the fourteen-year national decline in homophobia; the three 2020 experiments, on 6,869 voters, found d = 0.08 for exclusionary attitudes toward immigrants and 0.08 and 0.04 in the transphobia replications, durable to four months. That is support for the mechanism's reality at the scale of a conversation, short of a conversion; what the chapter rests on it is durability, which the designs measure, with magnitude left where the studies leave it.

The second sets the single paradigm inside its base. The meta-analytic literature on intergroup contact (Pettigrew & Tropp, 2006) finds prejudice reliably reduced by contact across hundreds of studies, with the largest effects where the contact is personal and sustained. That is the population-level trace of the same mechanism, and the reason the chapter treats canvassing as one instrument of deep persuasion rather than its proof.

The third concerns the base itself. Paluck, Green, and Green's (2019) re-analysis of the contact literature confines the estimate to experimental designs and finds the effects smaller there than the correlational record suggests. Few studies measure outcomes at a delay, and fewer still among adults on the prejudices that matter most politically.

The framework's stake accordingly rests where the durable measurements are: on the canvassing experiments above, whose delayed outcomes were the point of their design, and on whatever the contact base retains once the re-analysis has done its work. A base that thinned further under such scrutiny would narrow the evidence for the mechanism without touching the derivation. The Register's stake on Chapter 15 (§4.4) is to be read at the size this paragraph fixes.

Dryzek (2000) develops a broader conception of deliberative democracy, extending beyond formal polling to include informal public deliberation, social movements, and the discursive contestation of norms in civil society. Dryzek argues that deliberative democracy is not a single institutional form but a family of practices unified by the commitment to decision-making through the exchange of reasons rather than through aggregation of fixed preferences.

The contribution to Dryzek's project is the evaluative architecture. The "reasons" Dryzek identifies as the currency of deliberation are, on the present account, emotionally meaningful inputs that engage the evaluative modality and can produce real shifts in evaluative response. A reason that engages no evaluative response is inert — the field with no body in it of Chapter 14 (§A.5). It may be logically compelling and motivationally empty.

7 Objections and Replies

7.1 The utopian objection

Objection. The strict definition of peace is unattainable. Deep persuasion is slow, difficult, and frequently impossible.

Reply. The chapter does not claim that deep persuasion is always achievable or that enforcement and shallow persuasion are never appropriate. It claims that deep persuasion is the only mode that produces peace. Real societies operate with a mix of all three modes on different timescales. The strict definition serves a diagnostic function: it tells us what peace actually requires, which in turn tells us what the alternatives can and cannot achieve.

7.2 The agonistic democracy objection

Objection. Mouffe (2000, 2013) argues that political life requires ongoing disagreement — that the attempt to eliminate all conflict is depoliticisation.

Reply. Mouffe (2005) identifies a real feature of democratic life. She argues that the liberal attempt to achieve rational consensus,

paradigmatically Habermas's discourse ethics, is not merely utopian but dangerous, depoliticising fundamental disagreements that reflect irreconcilable differences in values and identity. The task of democratic politics is instead to transform antagonism, enmity, into agonism, adversarial respect, in which opponents are adversaries whose right to disagree is acknowledged rather than enemies to be destroyed.

The response runs along the line already drawn: Mouffe's agonistic transformation divides at §2.1 and §3.2. Strategic acceptance of the democratic framework is shallow persuasion at a high level of sophistication, reverting as the balance shifts. A shared preference for the adversary's right to disagree is deep persuasion at the framework level — level-zero contest housed inside level-one alignment — with durability the empirical discriminator, as everywhere in the taxonomy. Either reading records a real achievement, divided by the same line as the constitutional pluralism of §3.2.

The analysis disagrees with Mouffe on two points. First, on the claim that all fundamental political disagreements are irreconcilable. Some are, and agonistic management is the best available response; others are susceptible to deep persuasion through the kind of sustained evaluative engagement the chapter describes. The analysis does not claim that deep persuasion is always possible. It claims that deep persuasion is the only mode producing structural peace, and that Mouffe's agonism is the correct fallback when deep persuasion fails or is not yet achievable.

Second, on the claim that the preservation of disagreement is intrinsically valuable. Mouffe holds that democratic vitality requires ongoing conflict; the analysis holds that conflict is a condition to be resolved where possible and managed where not, and that the institutional apparatus should be preserved because disagreement is always liable to recur rather than because it is valuable.

7.3 The pluralism objection

Objection. Stable pluralism — a society in which groups with fundamentally different values coexist under a shared constitutional

framework — is a real political achievement that the chapter's binary treatment misses (Rawls, 1993; I. Berlin, 1969).

Reply. The chapter agrees that stable pluralism is a real and important achievement, and the account's diagnostic value is that it distinguishes the two achievements "stable pluralism" names together, along the line of §§2.1 and 3.2. A shared framework-preference is peace in the strict sense at level one; strategic acceptance is shallow persuasion at its most sophisticated, its instability latent. The binary treatment the objection alleges is what the levelled analysis replaces.

7.4 The moral limits objection

Objection. Should we deeply persuade a Nazi rather than enforce against him? Some preferences are morally monstrous.

Reply. The hierarchical ordering makes clear that enforcement addresses the immediate threat (survival) and shallow persuasion contains the ongoing danger (security). Deep persuasion, if possible, eliminates the preference that generated the danger (alignment). A Nazi who is imprisoned but not persuaded will resume his project if released. A Nazi who is deeply persuaded — who comes to feel the suffering his ideology causes — is no longer a Nazi. Enforcement manages the threat; deep persuasion alone eliminates it. Whether deep persuasion is possible in a given case is empirical; that it is the only mode achieving structural resolution is structural.

The empirical record counsels the sobriety the reply's conditional already encodes. The literature on exit from violent extremism finds disengagement — behavioural departure, a shallow-persuasion outcome sustained by changed incentives and circumstances — far more commonly than deradicalisation proper, the transformation of the underlying commitments, and finds the latter slow, relational, and comparatively rare (J. Horgan, 2009; Bjørgo & Horgan, 2008).

The taxonomy predicts that asymmetry, since transformation requires the emotionally meaningful engagement of the subject's own evaluative modality, which cannot be administered as a programme, while behaviour yields to management. That the deep mode is rare

where it is most needed is not an objection to the hierarchy but the hierarchy's own forecast — and the reason enforcement and containment remain, for the Nazi as for anyone dangerous, the modes the analysis itself prescribes in the meantime.

7.5 The incommensurability objection

Objection. Some disagreements involve incommensurable values (I. Berlin, 1969), so deep persuasion may be impossible where evaluative hierarchies are fundamentally different in kind.

Reply. The objection identifies a real structural limit. Where incommensurability obtains at level zero, the first resort is ascent rather than management: framework alignment, deep persuasion at level one about the terms under which the incommensurable values coexist (§2.1), remains available where value-alignment is not. Where even that is unavailable, shallow persuasion at its most sophisticated — strategically accepted constitutional arrangements, federal compromises, prudential agonism — is the best remaining outcome. The analysis still serves a diagnostic function, since it identifies which conflicts are resolvable at level zero, which at a higher level, and which require permanent management.

7.6 The manipulation objection

Objection. If deep persuasion transforms evaluative hierarchies, how is it distinguished from manipulation?

Reply. Deep persuasion works through the engagement of the modality with emotionally meaningful content; manipulation works through its bypassing or exploitation, with responses triggered by deception, selective framing, or cognitive bias, without engagement with the content. In the terms of Chapter 14 this is the line between a fitting response and an unfitting one: deep persuasion brings the agent to register the relevant features more accurately, above all the states of those her conduct affects, whereas manipulation engineers a response that misreads them (Chapter 14, §6.4). The line is close to Noggle's (1996), on which manipulation is the attempt to lead another astray, inducing beliefs, desires, or emotions that fall short of the ideals

governing them; the account takes that shape and fixes the ideal by the anchor, the accurate and complete registration of the affected parties, where Noggle leaves it to the manipulator's own standards.

The boundary can be unclear in specific cases while the distinction is sharp — and inherited rather than invented here. Chapter 15 (§4.4) states it as the discipline on enrichment: added value must bottom out in real registrations. The time-indexed transformation test (Chapter 15, §3.2) supplies the audit, and the standard it applies is the informed-registration anchor (Chapter 14, §6.5). A preference transformed by deep persuasion carries its history — the encounters and registrations that restructured it, traceable in principle — whereas a manipulated shift pleads a history keyed to registrations that were concealed, fabricated, or misread. This is why the fixed-point asymmetry below holds: the manipulation's sustaining misreading is itself the real registration its history lacks. Sunstein's (2015) nudging occupies the boundary, since a nudge that engages the agent's evaluative capacity by presenting consequences vividly is closer to deep persuasion, while one that bypasses it by exploiting default biases is closer to manipulation.

A deeper structural difference lies behind this difference in mechanism: manipulation has no stable fixed point. It works by inducing a response keyed to an incomplete or inaccurate picture of the relevant evaluative differences, so its product survives only while the gap between that picture and reality holds — the same contingent overrider §2 identified in every state on the conflict side of the ledger. Once reality manifests the features the manipulation concealed, the induced response loses its support, so the original preference, still untransformed, re-expresses itself. Manipulation is in this respect a species of shallow persuasion, reverting whenever its sustaining misreading is exposed.

Deep persuasion leaves nothing of the kind to expose, because the agent has already registered the relevant differences accurately, so no concealed reality remains whose arrival could undo the shift. None of this guarantees that a given manipulation will unravel, since reality may stay hidden, or the misreading may be renewed faster than it decays, and

it locates the asymmetry. Deep persuasion is stable because it is fitting; manipulation is unstable because it fits nothing that reality will confirm.

8 Conclusion

Deep persuasion is the only mode that produces peace. This is a strict finding. Enforcement stops the harm. Shallow persuasion prevents its recurrence. Deep persuasion eliminates the preference that generated it. The three modes are not alternatives among which a society chooses; they are layers through which a society progresses, each creating the conditions for the next.

The chapter yields three results, all constructed.

Three modes of conflict resolution. Conflict resolution admits of exactly three modes: enforcement (capacity-removal), shallow persuasion (preference-override), and deep persuasion (preference-transformation). The taxonomy is exhaustive because a preference can be addressed at exactly three levels of the evaluative architecture, and the third level's mechanism is inherited — deep persuasion is field-enrichment (Chapter 15, §4.4) between agents, carrying that result's discipline and its time-indexed audit. There is no fourth level.

The hierarchy of modes. The three modes stand in a hierarchical relationship grounded in structural dependency: enforcement addresses survival, shallow persuasion addresses security, and deep persuasion addresses alignment. The trajectory is progressive, since each instance of deep persuasion expands shared understanding, without being guaranteed, and the institutional apparatus must be preserved as a standing capacity.

Deep persuasion and structural peace. Deep persuasion is the only mode that produces structural peace, the state of maximal social freedom in which alignment dissolves the opposed preferences that are the sole social constraint on action. Lawmaking and law enforcement are distinct operations with different normative requirements. The derivation converges with Habermas's ideal speech situation while being grounded in the metaphysics of evaluative difference rather than

the pragmatics of communication, its conditional surviving the refusal to argue intact (§6.1).

If deep persuasion is the terminal mode, what conditions must a society satisfy for deep persuasion to be structurally possible? And once those conditions are met, what is the concrete mechanism by which private hierarchies become shared law? The next chapter derives the conditions, identifies the mechanism, and shows that the institutional form that preserves both is democracy.

Grades and Dependencies

1. Stance and method. All three results are *constructed*, on Chapters 15 and 16. The chapter's stance is on record (§1): descriptive, not prescriptive — its claims are hypothetical imperatives that command nothing of those the persuasion has not reached, a symmetry stated as plainly against the framework as for it (§6.1). Its method is on record too (§3.1): the informed-registration audit runs prospectively on three axes — concealment vs exposure, local reversion, generalisation — with verdicts provisional and durability the criterion they proxy. The canvassing evidence is weighed at its size: one instrument of deep persuasion, not its proof (§6.2). The Premise Register carries the chain.

References

Arendt, H. (1970). *On violence*. Harcourt, Brace & World.

Axelrod, R. (1984). *The evolution of cooperation*. Basic Books.

Bar-Tal, D. (2013). *Intractable conflicts: Socio-psychological foundations and dynamics*. Cambridge University Press.

Berlin, I. (1969). *Four essays on liberty*. Oxford University Press.

Bjørgo, T., & Horgan, J. (Eds.). (2008). *Leaving terrorism behind: Individual and collective disengagement*. Routledge.

Broockman, D., & Kalla, J. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, 352(6282), 220–224.

Burton, J. W. (1990). *Conflict: Resolution and provention*. St. Martin's Press.

Deutsch, M. (1973). *The resolution of conflict: Constructive and destructive processes*. Yale University Press.

Dryzek, J. S. (2000). *Deliberative democracy and beyond*. Oxford University Press.

Fishkin, J. S. (2009). *When the people speak: Deliberative democracy and public consultation*. Oxford University Press.

Galtung, J. (1969). Violence, peace, and peace research. *Journal of Peace Research*, 6(3), 167–191.

Habermas, J. (1990a). Discourse ethics: Notes on a program of philosophical justification. In *Moral consciousness and communicative action* (C. Lenhardt & S. W. Nicholsen, Trans., pp. 43–115). MIT Press.

Habermas, J. (1990b). *Moral consciousness and communicative action*. MIT Press.

Habermas, J. (1996). *Between facts and norms*. MIT Press.

Hart, H. L. A. (1961). *The concept of law*. Clarendon Press.

Hirschman, A. O. (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*. Harvard University Press.

Horgan, J. (2009). *Walking away from terrorism: Accounts of disengagement from radical and extremist movements*. Routledge.

Kalla, J. L., & Broockman, D. E. (2020). Reducing exclusionary attitudes through interpersonal conversation: Evidence from three field experiments. *American Political Science Review*, 114(2), 410–425.

Kelman, H. C. (1958). Compliance, identification, and internalization: Three processes of attitude change. *Journal of Conflict Resolution, 2*(1), 51–60.

Lederach, J. P. (2005). *The moral imagination: The art and soul of building peace*. Oxford University Press.

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 50(4), 370–396.

Mouffe, C. (2000). *The democratic paradox*. Verso.

Mouffe, C. (2005). *On the political*. Routledge.

Mouffe, C. (2013). *Agonistics: Thinking the world politically*. Verso.

Noggle, R. (1996). Manipulative actions: A conceptual and moral analysis. *American Philosophical Quarterly*, 33(1), 43–55.

Paluck, E. L., Green, S. A., & Green, D. P. (2019). The contact hypothesis re-evaluated. *Behavioural Public Policy*, 3(2), 129–158.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90(5), 751–783.

Pettit, P. (1997). *Republicanism: A theory of freedom and government*. Clarendon Press.

Rawls, J. (1993). *Political liberalism*. Columbia University Press.

Schelling, T. C. (1960). *The strategy of conflict*. Harvard University Press.

Sunstein, C. R. (2015). *Choosing not to choose: Understanding the value of choice*. Oxford University Press.

CHAPTER 22

Democracy and the Lawmaking Mechanism

Presupposes: the anchoring of moral disvalue in the affected party's response (Ch. 14, §6.4); the primary/added value taxonomy and field-enrichment (Ch. 15); the great moral conflation (Ch. 16); the perfection/infallibility distinction (Ch. 17); the three modes of conflict resolution (Ch. 21).

Establishes: the structural conditions for deep persuasion — expressibility, receptivity, and non-exclusion; equal deliberative standing, distinct from graded moral consideration; the lawmaking mechanism that converts private evaluative hierarchies into shared law, with its democratic identification; the limit on speech the three conditions draw; the diagnosis of bipolar authoritarianism.[1]

In brief. "Who are you to tell me what is right and wrong?" A tyrant asks it to silence a community; a relativist asks it to forbid persuasion; an absolutist answers it by placing the law beyond human reach. All three share one defect — authority-dependence — and democracy rejects the demand itself: lawmaking needs opportunity, not authority. Deep persuasion at scale requires three things with necessity: perspectives must be expressible, received on their merits, and no perspective excluded in advance — free speech, open speech, and equal standing are the best available means, not the necessity itself. The mechanism: each hierarchy entering the field carries perspectives that can restructure the others, until shared preference becomes shared law — values first, rules second. Two laws define the form: all equal *under* the law, all equal *above* it. Only democracy keeps both.

---

1 Introduction

If deep persuasion is the only mode that produces structural peace, what must a society look like for deep persuasion to be possible, and what happens when the conditions are met? The derivation proceeds in two layers that must be kept apart, because they carry different modal weight.

The structural layer. The first layer follows with necessity from what deep persuasion is. Deep persuasion transforms a preference by bringing the agent to encounter the perspective that would transform it.

This requires that the perspective be expressible, so it can enter the evaluative field; that it be received on its merits, so the listener engages its content rather than the speaker's identity; and that no relevant perspective be excluded in advance, so every hierarchy capable of transforming the field is admitted to it. Expressibility, receptivity, and non-exclusion are the conditions without which the mechanism cannot operate, and their necessity is of the cheap kind: they follow from what deep persuasion is, so the argument claims analytic necessity for these three and for nothing beyond them; the chapter's substance lies in the institutional layer.

The institutional layer. The second layer is institutional. Free speech, open speech, and democratic equality are the arrangements by which a society secures the three structural conditions: free speech protects expressibility, the cultivated disposition of open speech protects receptivity, and equal standing protects non-exclusion.

That these institutions are the best available means of securing the conditions is an empirical claim about how communication and prejudice actually work rather than a further entailment of the derivation, since a society that secured expressibility, receptivity, and non-exclusion by other means would satisfy the structural requirement equally. Keeping the layers apart locates what is necessary, the conditions, and what is contingent, their institutional realisation, and the distinction is observed throughout the chapter.

Non-exclusion carries the heaviest load, because it raises the question of who must be admitted to the field, which cannot be settled by stipulation. Any criterion for standing other than the capacity for evaluative engagement — race, creed, intelligence, culture — introduces false negatives, excluding beings whose perspectives could transform the field. The capacity to feel is the only criterion that excludes no being whose condition bears on the evaluative differences at stake. Section 4 develops this, and distinguishes the equal deliberative standing the mechanism requires from the graded moral consideration that empathic intensity tracks.

The argument derives the structural conditions for deep persuasion at scale — expressibility and receptivity — together with the institutions that secure them, free speech (§2) and open speech (§3), then grounds the third condition, non-exclusion, in equal standing (§4). It develops the lawmaking mechanism as the concrete operation by which private hierarchies become shared standards (§5), and shows that democracy is the institutional form preserving both the conditions and the mechanism (§6). The derivation ends there. What it explains comes next: why relativism and absolutism both obstruct democratic peace through a shared structure of authority-dependence (§7). The literature is engaged (§8) and the objections met (§9); the democratic failures, read as failures of the conditions, together with the institutional implications, are drawn out in an appendix (§§A.1–A.2), with the objection aimed at them (§A.3).

2 Free Speech as a Structural Condition

Deep persuasion requires communication. An agent whose evaluative hierarchy is to be transformed must first encounter the perspective that will transform it, the new secondary objects that will alter the added-value profile of her conflicting preference. Without communication these new secondary objects cannot enter the evaluative field, so the hierarchy cannot shift.

The two layers of §1 apply here first. The structural condition is expressibility: the perspective that would transform a hierarchy must be able to enter the field. The free-speech right is the institution that secures it under real conditions, where the power to silence is widely held and cheaply exercised. Necessity attaches to the condition; the right is the best available means; the arguments that follow are arguments about means. A society that guaranteed expressibility without a formal right would meet the condition, and none is known to do so, which is why the institution is needed.

Pettit (1997) argues that freedom is best understood as non-domination, the absence of arbitrary power over one's choices, rather than as non-interference, the liberal view. Free speech, on Pettit's

republican account, is valuable not merely because it enables the exchange of ideas but because its suppression constitutes domination: the arbitrary power of one group to silence another.

The framework converges with Pettit while providing a different grounding, since free speech is valuable because it is the structural condition for deep persuasion, and the suppression of speech disables deep persuasion in the affected domain. Pettit's non-domination and the present analysis's deep persuasion converge on the same practical requirement, free speech, for different reasons: non-domination targets arbitrary power, while deep persuasion targets the conditions for preference-transformation.

Free speech is therefore the structural condition that makes persuasion possible. The derivation proceeds from the structural priority of deep persuasion rather than from a prior commitment to individual autonomy (Mill, 1859/1999), a theory of rights (Dworkin, 1977), or the instrumental value of a "marketplace of ideas" (Holmes, 1919). If deep persuasion is the only mode that achieves structural peace, and deep persuasion requires communication, then expressibility is a structural precondition for peace, and the free-speech right is the institution that secures it. The nearest ancestor is the self-government tradition (Meiklejohn, 1948), where free speech is justified by the requirements of collective decision; the derivation here descends a step further, to the mechanism that makes the deciding transformative, while Schauer's (1982) caution that no single value underwrites the whole practice is absorbed as the two-layer analysis predicts.

Defending free speech means protecting the expression of opinions that others may dislike or find offensive, not just those already accepted. The preferences most in need of transformation through deep persuasion are the ones we find most objectionable. A society that permits speech only when it is agreeable has disabled deep persuasion for the conflicts that matter most. The most principled form of free speech is the willingness to let people we detest say things we detest, because the alternative to hearing them is enforcing against them, and enforcement manages conflict without resolving it.

A limit applies, and the three conditions draw it. Speech serves as a substitute for force, so when it is used to incite violence or coercion it becomes a surrogate for force rather than an alternative to it, and the derivation does not protect speech that disables the persuasive process itself.

The same test reaches the harder case the silencing literature presses (Langton, 1993; Maitra & McGowan, 2012). Speech that removes another party's expressibility, receptivity, or standing — that makes her testimony unspeakable, her hearers unable to take it up, or her perspective one the field will not admit — destroys a condition of deep persuasion for someone, so it falls outside the protection the conditions confer, for the same reason incitement does. The limit is therefore internal to the mechanism rather than a balance struck from outside it. Speech is protected as the medium of deep persuasion, so speech that disables that medium for another has no claim on the protection. Where the line falls in a given case is an empirical question about effects on expressibility, receptivity, and standing, where the conditions say what must be shown.

3 Open Speech as a Structural Condition

Freedom of speech alone is not enough. Deep persuasion requires not only a speaker who communicates but a listener willing to engage. Open speech is the listener's contribution to the persuasive process: the disposition to evaluate a claim on its merits rather than on the identity of the speaker.

The same two layers apply. The structural condition is receptivity: the perspective, once expressed, must be weighed on its content, and open speech names the disposition that realises it. Here the institutional layer is thinner, because a disposition cannot be mandated the way a right can be guaranteed. Institutions can protect the conditions under which open-mindedness forms — education, norms of debate, protection from reprisal for changing one's mind — while the receptivity itself is secured more by culture than by law.

Its opposite is close-mindedness: the dismissal of a claim based on prejudice against its source rather than assessment of its content. A person who assumes a child cannot understand complex mathematics and rejects the child's correct solution without examining it is close-minded. A person who assumes a non-believer cannot grasp moral truth and dismisses her moral argument without considering it is close-minded. In both cases, the speaker is assessed orthotically, checked against a standard of identity (age, creed, status), rather than engaged with on the substance of what she offers.

The connection to the great moral conflation of Chapter 16, the mistaking of conformity to a standard for empathic evaluation, is structural. Close-mindedness treats the speaker as a candidate to be checked against a standard of legitimacy: "Is she the right kind of person to make this claim?" Open-mindedness treats the speaker's claim as a contribution to be evaluated on its content: "Is what she is saying right?" When the question "who are you to tell me what is right?" replaces the question "is what you are saying right?", the orthotic assessment of the speaker has displaced evaluative engagement with the claim.

Both conditions are required. Free speech without openness is shouting into closed ears; openness without free speech is a willingness to listen that has nothing to hear.

What institutions can do for receptivity is limited and should be stated with the limit. Three instruments have evidence behind them.

Deliberative minipublics structure the encounter itself, since balanced information, moderated exchange, and time produce the measured opinion shifts Fishkin's polling records (Fishkin, 2009), and what the design supplies is a setting in which weighing a claim on its merits is the path of least resistance. Deliberative civic education rehearses the disposition young, treating disagreement as material for reciprocal reason-giving (Gutmann & Thompson, 2004). And norm-protection lowers the price of the disposition's most costly exercise, since a person who expects changing her mind to be read as weakness will not change it in public, so institutions that honour revision — in

legislatures, in science, in journalism — protect the conduct receptivity issues in.

The limit is that none of the three manufactures the disposition. Each lowers its costs, models its practice, or removes its penalties, while a population unwilling to weigh claims on their merits defeats them all. Receptivity is accordingly the condition of the three that institutions secure least and culture must carry furthest — a finding, since it locates where a democracy's health depends on what its laws cannot reach.

4 Democratic Equality

4.1 The grounding: relational value

Both free speech and open speech presuppose a form of social equality: the recognition that all persons have equal standing in the legislative process. This social equality is distinct from, indeed the near-inverse of, the equality of the diversity/equality partition (Chapter 3). The foundational term names a state without any internal distinction, whereas the equal standing invoked here holds among persons who are maximally differentiated; it is grounded, as this section shows, in relational value rather than in the metaphysical terminus. What grounds this recognition?

Two layers, in order. The analysis's answer has two layers, and keeping them in order is the section's work. The ground of moral standing is the being's own evaluative modality: her capacity, actual or potential, for states that are better or worse for her, so standing follows the feeling that is there rather than the feeling it occasions (§4.2).

The relational layer sits above the ground. Empathic response — her condition engaging our evaluative modality, ours engaging hers — is how standing is recognised, weighed, and made effective among persons, where the equality this section defends is a property of those relations. The onlooker's response detects and protects standing; what it never does is confer it, and the paragraphs that follow show why any onlooker-grounding begs the question the criterion must answer.

Young (2000) argues that democratic inclusion requires more than formal equality of participation, since structural inequalities in resources, cultural recognition, and institutional access prevent some groups from participating effectively even with formal standing. Her structural injustice is, on the present analysis, shallow persuasion masquerading as deep: the formal conditions for exchange are met, since everyone can speak, while the substantive conditions are not, since some voices carry less weight. The framework agrees that formal equality is insufficient, because deep persuasion requires the substantive capacity to be heard as well as the right to speak.

The adaptive preferences analysis of Chapter 15 (§6.6) identifies a further barrier to the open speech condition that Young's account of structural injustice does not fully capture. The problem is not only that dominated groups lack the formal right to speak, or even that they lack the substantive capacity to be heard. The deeper problem is that domination can shape the dominated group's own evaluative hierarchy so that they do not recognise their preferences as adaptive: they experience their subordination as natural, their silence as appropriate, their exclusion as deserved. Chapter 15 showed that adaptive preferences carry a specific structural profile, since the subject's primary-value response has been suppressed or overridden by added value — the consequential and derived value imposed by the oppressive structure.

When this mechanism operates at the social scale, the open speech condition faces a barrier the formal analysis alone does not reveal. A speaker who has been shaped by domination to regard her own perspective as worthless will not exercise the right to speak even when it is formally guaranteed. The evaluative orientation that would make speaking feel worthwhile has been suppressed by the structure the speech is meant to challenge.

Hermeneutical injustice. Fricker's (2007) concept of hermeneutical injustice names the epistemic dimension of this barrier. A group suffers hermeneutical injustice when it lacks the interpretive resources to understand its own experience, when the concepts, categories, and narratives available to it do not include the vocabulary

needed to identify its subordination as subordination rather than as the natural order. On the analysis, hermeneutical injustice is a specific form of adaptive preference operating at the epistemic level. The group's evaluative hierarchy has been shaped so that the derived value of the dominant narrative overrides the primary-value response to the group's own condition.

The open speech condition therefore requires more than institutional protections (free speech) and cultural dispositions (open speech). It requires the prior work of what might be called evaluative liberation: the process by which dominated groups recover access to their own primary-value responses, recognise their preferences as adaptive, and constitute the evaluative orientation that makes participation in deep persuasion possible.

This is itself a form of deep persuasion, directed at the dominated group's own hierarchy rather than at the oppressor's, with the aim of restoring the primary-value responses that oppressive structures have suppressed. Dotson's (2014) concept of epistemic oppression, the systematic exclusion of knowers from knowledge-producing practices, reinforces the point. The conditions for deep persuasion include the epistemic and evaluative capacity to recognise that one has something worth saying, as well as the institutional opportunity to speak.

This relational grounding converges with Anderson's (1999) democratic egalitarianism and Scheffler's (2003) relational account of equality, since both hold that equality concerns the relationships among persons — mutual respect and equal standing — rather than the levelling of distributions. The account agrees, and adds the explanation of why the relationship takes this form. Evaluative differences are constituted by the modality's response, and empathic evaluation, emotion directed at emotion, is how the moral standing others hold in their own right is recognised and weighed.

A person in a vegetative coma may have severely diminished evaluative capacity, while still triggering real empathic response in others: her mother at the bedside singing to her, a nurse adjusting her

position to prevent discomfort. The empathic response is real; it is not what gives her standing.

She has standing in her own right, as a being whose evaluative modality, diminished or suspended, remains hers, retaining the capacity, actual or potential, for states that are better or worse for her. The empathy at her bedside detects and protects that standing rather than conferring it. Only where such capacity is determinately and permanently absent does the criterion run out, and there the care of others honours the person who was rather than constituting a standing that is.

The diagnostic. A justification for this criterion can be stated as a diagnostic. The question "what criterion determines who gets standing in the legislative process?" must be answered without circularity. Not by asking which beings move us, which would make the onlooker's response the ground and beg the question, but by asking which beings have an evaluative modality of their own: the capacity to register what is done to them as better or worse for themselves.

Run the alternatives. If standing is grounded in race, other races are excluded. If in intelligence, the less intelligent are excluded. If in culture, other cultures are excluded. If in DNA, edge cases are excluded. Any specific feature other than the capacity to feel, or the potential to feel, risks excluding beings that have valenced states of their own and so possess moral standing whether or not they happen to move an observer.

The capacity for emotion is what every being with standing shares. Infants, the cognitively impaired, the comatose (potential feeling), and the developing foetus (potential feeling) all have valenced states of their own, present or potential, for an act to be good or bad for. It is that capacity, rather than the empathy they happen to elicit, that grounds their standing, so no narrower criterion can be substituted without wrongly excluding some of them.

The cost is imperfect specificity, since the criterion also includes animals, who do feel. But this is a feature rather than a defect, because animals should receive moral consideration: they have valenced states

of their own. The hierarchy of consideration tracks what is at stake in a being's own registrations, which the intensity of empathic response tracks fallibly, so animals receive real moral consideration without receiving equal legislative standing with humans.

Why standing among humans is equal. A gap in this reasoning must be closed, because the grounding just given is graded while the conclusion it is meant to support is not. If standing tracks the intensity of empathic response, and intensity comes in degrees, as the animal case concedes, then nothing yet explains why standing among humans is equal rather than proportioned to the intensity each person elicits.

The equality does not come from the relational grounding; it comes from the mechanism. What the lawmaking mechanism requires of a participant is the capacity to introduce a perspective that can transform the evaluative field — novel secondary objects the other hierarchies have not registered. This capacity is not graded in the way empathic intensity is graded, since a perspective either can enter the field as a candidate for uptake or it cannot, and every being with an evaluative hierarchy of its own carries one. Perspectives differ in transformative power; eligibility for uptake is binary, the field's response once a perspective has entered is graded — which is where Landemore's diversity effect lives — and equal deliberative standing is a claim about the first, the argument against weighting the second being the incentive argument of §6.1. For the purpose of deep persuasion, then, each such being counts equally as a source of perspectives, however much the moral consideration owed to it varies with what is at stake for it.

Two distinct things have thus been grounded in two distinct ways. Moral consideration is grounded in the capacity to feel and graded by what is at stake for the being, which empathic intensity tracks fallibly (§6.2), and so it extends to animals in degrees. Equal deliberative standing is grounded in the capacity to contribute a transformable perspective and is not graded, which is why it holds equally among all who possess an evaluative hierarchy. Equal standing among humans is therefore neither an additional commitment bolted onto the derivation nor an inference from a graded property to an ungraded conclusion; it follows from the binary structure of what the mechanism requires.

A scope note is required, because the criterion touches the hardest cases. That a being has standing — actual or potential capacity to feel — is not, by itself, a policy conclusion. Where standings conflict, as they do most acutely in the abortion debate, resolution requires two further tasks this chapter does not attempt: the empirical determination of when the capacity to feel obtains, which the framework assigns to evidence (§4.2), and the weighing of one being's standing against another's where both are engaged. The criterion fixes what is at stake in such conflicts, whose condition counts, and deliberately fixes nothing else, so the chapter's claims about deliberative standing and graded consideration are inputs to those debates rather than verdicts upon them.

The potentiality clause requires one delimitation, and the standard one serves. What counts is identity-preserving development toward the capacity to feel. The foetus is a continuant on such a trajectory; a gamete is not, since nothing develops from gamete to feeler without a constitution-changing fusion that ends the gamete's own career. The clause therefore does not slide down to every cell from which a feeler could causally arise, because it tracks the individual that would do the feeling.

The criterion's empirical framing covers the novel case at the other end. Whether any artificial system possesses or approaches the capacity to feel is a question of evidence (§4.2) rather than of substrate, and standing follows the capacity wherever it is found. That is a virtue of a criterion drawn at the level of feeling rather than species.

The clause, so delimited, still takes a side where sides are contested, and the ground for taking it should be argued rather than assumed. Two rivals fix the cost of the choice.

A criterion of actual capacity only — standing for present feelers alone — avoids the potentiality debates at the cost of false negatives of just the kind the criterion was built to forbid. The reversibly comatose patient, whose modality is suspended rather than absent, would hold standing only as the object of others' empathy, which is the onlooker-grounding the section has just rejected. A backward-looking amendment — actual capacity now or previously — recovers her, though only by

conceding the operative principle: it grounds present standing in the individual's own trajectory through time. Once trajectory is admitted as standing-relevant, the asymmetry between "has felt" and "will, by identity-preserving development, feel" needs a principled basis rather than a stipulated one.

The framework locates the basis in what standing is for. Standing follows the being for whom things can be better or worse, and a continuant on an identity-preserving path to feeling is a determinate individual to whom the arrival, or the prevention, of that capacity happens. There is therefore a subject of the loss, which is what the gamete case lacks (Marquis, 1989, builds a rival deprivation argument on the loss the clause makes visible; McMahan, 2002, presses the identity questions hardest against it). The weighing tradition (Thomson, 1971) begins where this section stops.

None of this converts the clause into a verdict on the contested cases. The criterion fixes whose states count rather than how conflicting standings weigh — the pregnant woman's standing is as fully grounded as any — and the weighing is where the substantive work of the abortion debate lies. What the argument does fix is the location of the disagreement. A critic who resolves the identity questions differently, holding with McMahan that the early foetus is not yet the individual who would later feel, rejects the clause at the metaphysics of the continuant rather than at the criterion of feeling. The framework's claim is that this, and nowhere shallower, is where the dispute belongs.

4.2 Empathy, pseudo-empathy, and the boundary of standing

Empathy is the evaluative modality responding to the actual or potential emotional states of another being. But emotional responses can also be directed at objects that do not feel. A child feels protective love for a teddy bear; an adult feels attached to a childhood home. These responses are real as feelings in the responder, while the object has no emotional states for the response to be directed at. This is pseudo-empathy: a response with the phenomenological character of empathy and without the foundation, since the other's condition does not include emotional states.

The boundary of moral standing tracks the boundary of feeling — of beings with valenced states of their own, rather than of beings that happen to elicit feeling in us. The framework identifies the question any debate about standing must answer: does this subject feel? It is a question about the subject's own states, so a being can hold standing that no onlooker yet answers to, and an object can draw tender responses while holding none. Standing follows the feeling that is there, not the feeling it occasions.

The question is empirical, about actual or potential emotional states, rather than metaphysical. Disputes about where the boundary falls — early-stage foetuses, severely brain-damaged patients, advanced AI systems — are empirical disputes about who feels, and the framework predicts that they will be resolved through the accumulation of evidence bearing on this question. The criterion is the one the metaethics anchors value in, registration, actual or potential, corrected only for accuracy (Chapter 14, §§6.4–6.5), which is why standing and value share a boundary and a substrate-indifference: both follow the feeling wherever it is found.

4.3 Animals and the hierarchy of standing

Animals. Animals feel, so empathic response to a suffering animal is empathy: the animal's emotional states are real. The question of why animals are not typically granted equal legislative standing with humans is not a question about whether they feel, since they do, but about what is at stake in their registrations and how their perspectives enter the field. The hierarchy of consideration tracks the stakes, which empathic intensity registers fallibly, rather than a binary line between "feels" and "doesn't feel". The claim concerns the weighting of consideration among beings whose standing the capacity criterion has already fixed (§§4.1–4.2), not the ground of standing itself.

This is a substantive metaethical claim that should be stated against the major alternatives. Singer (1975) argues for equal consideration of interests regardless of species, a utilitarian position that rejects the intensity-tracking the framework identifies. Korsgaard (2018) develops a Kantian case for animal moral standing grounded in animals' status as

ends in themselves. Nussbaum (2006) applies the capabilities approach, arguing that animals possess capabilities — for pleasure, movement, social affiliation — that ground moral claims.

The analysis's position is descriptive as well as structural, since consideration does track empathic intensity in actual moral practice, and the framework explains why. The normative standard is the stakes themselves, and the practice answers to them as any registration answers to its object (Chapter 14, §6.4): deeper empathic engagement with animals, through closer contact, greater understanding of their emotional lives, or cultural shifts in evaluative sensitivity, would bring the practice's intensity into line with what is at stake, expanding the circle of consideration. This is consistent with the progressive expansion that deep persuasion makes possible.

One thread in the two-track resolution of §4.1 must be pulled straight here, because the binary grounding of deliberative standing — every being with an evaluative hierarchy carries a perspective — appears to hand animals the equal legislative standing this subsection withholds. The reconciliation runs through the mechanism's own conditions. Contributing a perspective to the deliberative field is subject to expressibility (§2), since a perspective enters as a candidate for uptake only if it can be articulated into the field, answered, developed, and revised in exchange. Animal hierarchies are real and their registrations morally weighty, while they enter the field the only way they can: as registered subject matter, carried in by evidence and by the empathic uptake of participants who can articulate what the animal cannot.

The equality of deliberative standing therefore holds among contributors, beings whose perspectives can meet the field's give-and-take. Beings whose perspectives enter by proxy hold full moral consideration, graded as §4.1 grades it, together with a claim to faithful representation — a claim the expanding engagement described above makes progressively more demanding. This marks no deficit of feeling or of standing before the law's protections; it marks the difference between a perspective that argues and a condition that is argued for, where the mechanism needs both. The claim has a developed neighbour

in Donaldson and Kymlicka (2011), where membership rights for domesticated animals include representation in political decision; Regan (1983) grounds ungraded standing in subject-of-a-life status. The present account differs on the grading, since moral standing follows registration by degree while the deliberative claim is binary, so representation tracks contribution channels rather than species membership.

The line within humanity. The line runs through humanity as well, and the account should say so rather than leave the animal case to imply it. An infant, a person in late dementia, an adult who has never spoken: each has an evaluative hierarchy of her own, and each enters the deliberative field as the animal does, as registered subject matter carried by evidence and by the empathic uptake of those who can articulate what she cannot.

Her standing before the law and her claim on moral consideration are full and ungraded in the sense §4.1 fixes. What she holds in addition is the claim to faithful representation, and what she lacks is only the give-and-take that makes a perspective a contributor rather than a condition argued for. Equal deliberative standing among humans is therefore equal standing among contributors, with the proxy claim extending to every human whose perspective enters by another's voice — which is why the mechanism's health depends on the reach and accuracy of that representation as much as on the rights of those who speak for themselves.

Democratic equality is therefore a claim about standing in the legislative process rather than a sharp binary. All beings that trigger empathy have a claim to consideration, while the scope of that consideration is itself a subject of ongoing deep persuasion, conducted through the very mechanisms this chapter derives.

A note of epistemic humility belongs here. The empathy/pseudo-empathy distinction depends on the empirical question "does this subject feel?"; this is the question contested in debates about foetal consciousness, AI consciousness, and the moral status of patients in persistent vegetative states. The framework identifies the right question

without resolving the boundary disputes. Different answers to the empirical question will generate different scopes of moral standing, so the contribution is to specify what is at stake in the answer rather than to settle it.

4.4 The two laws of social equality

The first law: all are equal under the law. No individual or group is exempt from the standards the society has adopted. This principle is ancient and well-established: it is the foundation of rule-of-law traditions from Roman jurisprudence through the Magna Carta to modern constitutionalism.

The second law: all are equal above the law. No individual or group is excluded from shaping those standards. This principle requires more development, because its derivation in the present account differs from prior formulations of democratic participation and the difference has real consequences.

The existing literature grounds equal participation in autonomy (Kant), fairness (Rawls's original position), non-domination (Pettit), relational equality (Anderson), or the public realisation of equal advancement of interests (Christiano, 2008). Each derivation arrives at the same practical requirement, that citizens should have equal opportunity to shape the rules they live under, though each proceeds from a prior commitment to a value: autonomy, fairness, freedom, or relational respect.

The present account derives the second law from the lawmaking mechanism (§5). If deep persuasion operates by introducing new secondary objects into each agent's evaluative field, and if the mechanism's capacity to produce alignment depends on the range of evaluative hierarchies present in the deliberative field, then excluding any person from the field excludes the secondary objects her perspective would introduce. Each exclusion reduces the mechanism's capacity to restructure composites and produce shared preferences. The second law is therefore a structural requirement of the mechanism rather than a

moral aspiration, since it follows from how deep persuasion works rather than from a prior commitment to any particular value.

The derivation has a diagnostic consequence that prior formulations do not provide. The under/above pair constitutes a two-cell diagnostic that classifies political structures by which law or laws they satisfy:

Structure	**Equal under the law**	**Equal above the law**
Democracy	Yes	Yes
Authoritarianism	Yes (nominally)	No — foreclosed by power
Privilege	No	Yes (for the privileged)
Relativism	Yes	No — legislation impossible
Absolutism	Yes	No — legislation unreformable
Tyranny	No	No

Equal subjection without equal opportunity to shape the law is authoritarianism. Equal opportunity without equal subjection is privilege. Relativism upholds the first but disables the second (no one can legislate, because moral persuasion is deemed illegitimate). Absolutism upholds the first but violates the second (no human voice can alter the law). Only democracy, the conjunction of equal subjection and equal opportunity, preserves both. Violate either and democracy is structurally disabled; violate both and you have tyranny.

Where the argument stands: the three structural conditions have been derived and their institutional realisations identified (§§2–3), and non-exclusion has been grounded in a criterion of standing that no narrower feature can replace (§4). What follows is the operation those

conditions make possible — the mechanism by which private hierarchies become shared law.

5 The Lawmaking Mechanism

The preceding sections established who participates in the democratic process (§4) and under what conditions (§§2–3). This section addresses the prior question: what actually happens when the conditions are met? The answer is the lawmaking mechanism — the concrete operation by which deep persuasion transforms private evaluative hierarchies into shared public standards.

5.1 The mechanism stated

The mechanism operates through the evaluative architecture established in Chapter 15. Each agent enters the deliberative field carrying her evaluative hierarchy: her ordered ranking of objects by the intensity and orientation of the evaluative responses they elicit, constituted by the composite of primary and added value. Through sustained exchange, each agent encounters secondary objects she had not previously registered: the suffering caused by a policy she had supported, the consequences she had not considered, the experiences of people she had not encountered. These new secondary objects alter the added-value profile of her evaluation. The total value she assigns to the object of disagreement shifts. Her hierarchy restructures.

The process can be stated concretely. Suppose ten people are trying to agree on a set of behavioural rules, and initially they all disagree. To move toward agreement, each must engage with the value hierarchies of the others: is life more important than family? is health more valuable than wealth? does freedom outweigh convenience?

In practice this process requires each person to adjust the structure of others' evaluations — adding or removing elements, strengthening or softening emotional weight — until a shared preference emerges. This often involves reaching a more complete and refined empathic understanding of each other's primary and secondary objects of evaluation (Chapter 15).

A participant who initially values economic growth above environmental protection may, through sustained engagement with someone whose children suffer from pollution-related illness, come to feel the empathic weight of health consequences she had not previously registered. The added-value profile of her evaluation shifts, since the secondary objects associated with environmental regulation now include the suffering of identifiable persons, so the total value she assigns to environmental protection increases.

The new material has also reconstituted the object of evaluation itself. She is no longer evaluating "environmental regulation" but "environmental regulation whose absence causes identifiable children to suffer", and her primary response to the reconstituted object may differ from her primary response to the original (Chapter 15, §3.3). The mechanism therefore operates at both levels simultaneously, adding new evaluative weight to the composite and changing what is being evaluated. She has been neither overridden nor manipulated; she has been exposed to evaluative differences she had not previously encountered, and her modality has responded.

When enough hierarchies have shifted toward alignment on a specific preference, the shared preference is codified as law. Making law through persuasion is ultimately about sharing values first and rules second. The rules codify preferences that have been shaped through evaluative exchange.

Law is therefore practically necessary without being essential in principle. It is the institutional scaffolding the mechanism requires while alignment remains incomplete: the codification of partial convergence in a society whose evaluative hierarchies have not yet fully aligned. A society in which every preference had been aligned through deep persuasion would need neither law nor enforcement, because shared preferences would produce shared behaviour without external constraint. This condition corresponds to a full rotation of the clock of peace introduced in Chapter 21 (§4): the regulative ideal in which every instance of deep persuasion has expanded shared understanding until enforcement and shallow persuasion have become structurally redundant.

The clock is a regulative ideal, so no actual society may complete the rotation. But the lawmaking mechanism's trajectory points toward it, since each law the mechanism produces is a record of alignment achieved so far, and each instance of deep persuasion reduces the domain in which law is needed. The mechanism does not merely produce law; it works toward the condition in which its own product becomes unnecessary.

5.2 The mechanism's relationship to the evaluative architecture

The lawmaking mechanism is not an additional postulate. It is what the evaluative architecture of Chapters 14–15 does when operated under the conditions of Chapters 21–22. The modality constitutes oriented differences (Chapter 14); total value is a dynamic composite responsive to every new secondary object (Chapter 15); the subconscious jump delivers evaluative content involuntarily (Chapter 18); deep persuasion is the restructuring of the hierarchy by emotionally meaningful content (Chapter 21).

The lawmaking mechanism is this transformation operating at scale, under institutional conditions that preserve its integrity. Free speech ensures the secondary objects can be communicated; open speech ensures they are received; democratic equality ensures all relevant perspectives are available. The mechanism is the evaluative architecture in its political deployment. It is not a metaphor for how democracy works. It is a description of the operation that produces shared law from private hierarchies: the same operation that produces individual preference-change from the introduction of new secondary objects, now conducted among multiple agents simultaneously.

This is why the mechanism carries the weight of a derived result. Three things are already in hand: structural peace requires deep persuasion (Chapter 21); deep persuasion is the restructuring of an evaluative hierarchy by emotionally meaningful content (Chapters 14–15, 21); and that restructuring, conducted among agents, requires expressibility, receptivity, and non-exclusion (§§2–3).

Given these, the operation that converts private hierarchies into shared standards under conditions of disagreement follows by definition — it is what pursuing alignment under disagreement consists in — and the row that carries it in the Register is graded constructed for that reason. What stays open is the institutional embodiment: the claim that democracy, in the two-layer form developed below, is the political structure that houses the mechanism. Other arrangements could in principle realise the same operation.

5.3 Debate, voting, and the best available standard

Debate is the medium through which the mechanism operates. In debate, agents communicate the secondary objects that populate their evaluative fields — the experiences, consequences, and perspectives that have shaped their hierarchies. The exchange is not merely informational; it is evaluatively engaging, because the secondary objects communicated carry emotional weight that can restructure the listener's hierarchy.

Voting resolves a stalemate. A democratic vote means: if we cannot persuade each other, we must still act — peacefully. The vote prevents disagreement from escalating into violence. After the vote, persuasion may continue. What is lost at the ballot may yet be won in the future — through speech, not force. Voting is not an expression of relativistic equality among preferences; it is a procedural fallback when the mechanism has not yet produced sufficient alignment.

Waldron (1999) grounds the same fallback differently, and the difference is worth marking. For him, persistent disagreement is the circumstance of politics rather than a shortfall from alignment, and majority decision is what respects the disagreement rather than what waits out its removal. The present account treats the disagreement as a level-zero fact housed by a level-one preference (Chapter 21, §2.1), with the vote as what governs it meanwhile; Waldron's is the more austere reading, on which the housing is all there is. The divergence is over whether alignment is an aim or an illusion, and the clock of peace is what the framework stakes on the first.

Waldron's charge, at full weight. Waldron's case deserves its full weight before the divergence is costed, because it carries a charge against the aim itself. On his account (Waldron, 1999), majority decision is how equals respect one another under disagreement: each counts for one, none is asked to defer to a superior insight, and a theory that treats persisting disagreement as a shortfall — as matter awaiting transformation — has already graded some citizens' considered convictions as errors in waiting, which is its own failure of respect.

The reply runs through the levelled analysis of Chapter 21 (§2.1). What the framework marks for transformation is located wherever transformation can be had, and a level-zero conviction housed within a level-one alignment is graded resolved with respect to that alignment, with the conviction standing. The citizen who loses the vote is owed, on the framework's account, the standing conditions under which her persuasion continues at full weight (§6.1): the respect Waldron locates in the count, the framework extends past it.

What separates the views is a further premise Waldron carries and the framework declines: that the persistence of level-zero disagreement among the reasonable is permanent in principle across every level. The framework treats that as the open question the durability record bears on, since the federal and constitutional cases of Chapter 21 (§§2.1, 3.2) are level-one alignments with measured histories, so the dispute between the two accounts is placed where evidence can reach it.

Landemore (2013) provides the epistemological argument for this structure: cognitively diverse groups, deliberating under conditions of inclusion and mutual respect, systematically outperform homogeneous expert groups in problem-solving. Landemore's "democratic reason" is, on the structural analysis, the epistemic consequence of the mechanism operating at scale. When a cognitively diverse group engages in evaluative exchange, the range of secondary objects brought to bear on the problem exceeds what any individual or homogeneous group can provide, so the resulting assessment is more firmly aligned with the relevant evaluative differences. Cognitive diversity improves outcomes because it expands the evaluative field, introducing secondary objects,

alternative perspectives, and evaluative orientations that would otherwise be absent.

Within a culture of free and open-minded discussion, the proposal that best withstands scrutiny naturally rises to the top. This proposal is always the best available. But it is not absolutely the best, because a better alternative may emerge tomorrow — new secondary objects can always enter the field. Democracy is structurally open: it treats every adopted standard as the current best, subject to revision through future persuasion. This is the political expression of the epistemological truth that perfection does not entail infallibility.

Fishkin (2009) provides empirical illustration, though not confirmation, of the mechanism at work. His deliberative-polling findings — the significant, durable opinion change after structured engagement with opposing perspectives set out in Chapter 21 (§6.2) — bear directly here. On the present analysis, deliberative polling is the lawmaking mechanism in controlled institutional form, since the structured exposure to others' evaluative hierarchies provides the emotionally meaningful input that restructures the participants' composites and produces real shifts in evaluative response. Deliberative democracy's canonical statement (Cohen, 1989) grounds legitimacy in reasoned agreement; the present derivation grounds the institutions one level down, in the mode of change agreement rides on.

The findings are consistent with the structural claim and show that the mechanism can operate, while they do not establish it, because the claim concerns what deep persuasion requires rather than how often, or under what conditions, it succeeds. The structural result stands or falls on the derivation; the empirical work shows only that the derived mechanism is not merely notional.

6 Democracy as the Institutional Form

6.1 The democratic structure

Democracy, on the present account, is not a single device but the institutional form that does two structurally distinct things,

corresponding to two of the three modes of conflict resolution (Chapter 21, §3). The first is the deep layer: democracy institutionalises the standing conditions for ongoing deep persuasion — the expressibility, receptivity, and non-exclusion of §§2–4, realised as free speech, open speech, equal standing, and the deliberative fora in which hierarchies are exchanged. This layer is necessary in the strict sense, since a society that abandons it forecloses peace itself rather than merely one means to it.

The shallow layer. The second is the shallow layer. At any moment most conflicts remain unresolved by deep persuasion, and a society must still act on them, without waiting for an alignment that has not yet come and without surrendering the decision to whoever is strongest.

Majority voting is the procedure that decides this residue, and its justification is comparative rather than foundational. By the three-mode hierarchy of Chapter 21, an unresolved conflict can be met only by enforcement, by shallow persuasion, or by being left to fester. Majority voting is the least-bad shallow-persuasion instrument among these, because it settles the matter peacefully, binds the parties to no permanent verdict, and leaves the deep-persuasion channel open, so what is lost at the ballot may still be won by persuasion afterwards.

Voting is therefore optimal given the hierarchy, not necessary in the way the deep layer is. It does not produce peace; it manages the not-yet-pacified residue without violence while peace is still being made. A vote that overrides a persisting minority preference is, in the terms of Chapter 21 (§3.2), the strategic case — shallow persuasion — not the alignment the deep layer seeks.

The two-layer reading also fixes the framework's answer to the standing challenge to democratic equality itself. Brennan (2016) argues that if political outcomes were reliably improved by weighting votes toward the demonstrably competent, nothing about democracy's value would forbid it — the epistocratic proposal. The deep layer is untouched by the proposal and proof against it, since deliberative standing is binary, grounded in the capacity to contribute a transformable perspective (§4.1), while competence is a graded property of just the

kind the derivation showed cannot ground exclusion without costing the mechanism the perspectives it needs.

The shallow layer is where the proposal bites, and two considerations answer it there. First, the residue instrument decides just those conflicts deep persuasion has not yet reached, so weighting it by competence converts a temporary, contest-open settlement into a standing hierarchy of voice, telling the down-weighted that the channel still open to them, persuasion, now terminates in a decision where their uptake counts for less. That corrodes the very incentive structure the two-layer design protects, because what is lost at the ballot must remain winnable by persuasion; it remains so only if won persuasion counts at full weight.

Second, Estlund's (2008) qualified-acceptability requirement states in liberal vocabulary what the lawmaking mechanism entails in this one: any criterion of competence would itself have to be adopted as a shared standard, which means earning acceptance within the deliberative field it would then constrain. No such criterion has survived that field, since every candidate encodes some hierarchy's contestable conception of the political good. The epistocrat's counterfactual — a competence criterion beyond reasonable contest — describes a world in which deep persuasion has already aligned the relevant preferences, at which point the weighting is idle. Short of that world, the criterion is one faction's standard enforced on the rest, which is the second law violated in the counting room. Bell's (2015) political meritocracy, the Confucian form of the proposal, selects the competent by examination and performance, and its defence against Estlund's requirement is a tradition that has accepted the criterion; the reply is the one §9.1 gives the tradition, since acceptance sustained by the authority of the selecting hierarchy is shallow persuasion, and its durability under the withdrawal of that authority is the test.

Judicial review. Judicial review poses the two-layer question from the institutional side. Waldron (2006) argues against strong forms of it: where citizens disagree in good faith about rights, the disagreement belongs in the legislature, and handing it to courts substitutes one contested judgement for another while subtracting the equality of the

vote. The framework's account of evaluative volatility (Chapter 21, §5.2) pulls toward entrenchment and judicial independence, so the two accounts meet at a testable point.

The two-layer reading locates it. Buffering that protects the deep layer's standing conditions — the speech, standing, and revision channels through which what is lost at the ballot may still be won — is licensed by the derivation, since those conditions are what every future persuasion uses. Buffering that entrenches contested outcomes beyond the reach of persuasion has crossed into the unreformability Chapter 17 diagnoses, and on that side of the line the framework's verdict and Waldron's coincide. The dispute that remains is empirical at its centre: which institutional designs protect conditions, and which harden outcomes, is a question for the comparative record §A.2 assigns its predictions to.

Holding the two layers apart removes an equivocation the word "democracy" invites, since it names at once a deliberative ideal and a decision procedure. Run together, the two breed opposite errors. Treating the vote as though it produced alignment is the majoritarian error, since it reads a shallow-layer settlement as deep-layer peace, which is the tyranny of the majority diagnosed in §A.1. Treating deliberation as though it could dispense with any procedure for the residue is the utopian error, since it forgets that peace is made only gradually and that the unaligned remainder must be governed in the meantime. Democracy is the structure that maintains both layers at once, which is what distinguishes it from the authoritarian alternatives, examined in §7, that collapse one layer into the other.

6.2 Empathy and pseudo-empathy in democratic practice

Bloom (2016) argues that empathy is a poor guide to moral and political action, since it is biased, because we empathise more with those who look like us; innumerate, because we empathise more with identified individuals than with statistical populations; and manipulable, because skilled communicators can weaponise empathic responses. If the account holds that deep persuasion operates through empathic engagement, Bloom's critique poses a direct challenge: empathic

engagement may produce worse political outcomes than rational deliberation.

The response extends Chapter 16 (§6.6) to the political case. Bloom's critique targets what the framework identifies as pseudo-empathy, responses captured by bias, framing, and in-group preference, and the biases are distortions produced by the same mechanisms that produce adaptive preferences (Chapter 15, §6.6): the added value of in-group membership, the derived value of perceived similarity, and the consequential value of proximity overriding the primary empathic response. The corrective is therefore not to abandon empathy but to correct the distortions through the orthotic superstructure — the two-layer analysis in political application, with the empathic foundation providing the moral content and the superstructure the corrective mechanisms that protect empathy from its own distortions.

6.3 From truth to peace

Receptivity varies with confidence. Uncertainty opens a listener to suggestion; growing conviction can close her to perspectives that might improve on her own. The pattern invites a false generalisation — that stronger conviction entails greater dogmatism, so that a confident participant is already on the road to closure.

The two-dimensional classification of Chapter 17 separates the conviction from the closure. A certainty grounded on the logical axis and held open to challenge is compatible with full receptivity, since a confident idea earns its standing by surviving criticism, and a participant who defends her view inside a practice of open exchange is exercising the deliberative disposition, with her conviction as a contribution to the shared pursuit rather than an obstacle to it.

The connection between the best available standard and the pursuit of truth is structural; the argument for it is Mill's (1859/1999, ch. 2), restated in the account's vocabulary, which adds only the mechanism beneath the contestation, deep persuasion, and the reason it must be open to every hierarchy. If the best available standard is the one that has survived the most rigorous scrutiny, then the democratic process —

competing standards tested through free and open exchange — is the institutional mechanism best suited to identifying it. Democracy does not guarantee truth, while it preserves the conditions under which truth-approximation is possible: open contestation of standards by all affected parties, with no standard exempt from challenge. Its capacity to resolve disagreements through persuasion rather than force is what makes it the path toward peace.

The derivation ends here. What it explains follows: the authoritarian alternatives to democracy, diagnosed in §7 and, in the appendix, the failures democracies produce and the institutional forms that preserve or disable the conditions (§§A.1–A.2).

7 Relativism, Absolutism, and the Bipolar Dynamic

This section applies the derivation: it diagnoses why two influential alternatives to democracy — moral relativism and moral absolutism — obstruct peace through a shared structural defect. The diagnosis explains and does not establish; the Register carries it at its constructed tier.

7.1 The common structure: authority-dependence

The emergence of relativism and absolutism, and their shared structural defect, can be set out through one staged sequence. A small community deliberates over its moral preferences in turn.

A first speaker breaks the practice by force: “I am more powerful than all of you. Who are you to tell me what is right and wrong?” His will becomes the rule, and the community lives under tyranny until he dies. United by that experience, it writes its first law — all are equal under the law — and deliberation resumes. A second speaker then converts the remedy into a refusal: “Since we are all equal, who are you to tell me what is right and wrong?” The question carries the day, persuasion lapses as illegitimate, and no further rules form. The community has passed into relativism.

A third speaker restores legislation by removing it from human hands: “I have discovered a set of laws that does not come from me, but from a superior source.” Where the claim is believed, the community

lives under a code no one can challenge, since its presumed author stands beyond reach. It has passed into absolutism.

One clarification fixes the target before the diagnosis runs, because the name is shared with a position this is not: the target is a stance of political life whose adherents typically hold no metaethical view at all; the philosophical relativist is exempted by name below. Moral relativism as a metaethical thesis — that a moral claim's truth is relative to an agreement, a framework, or a standpoint (Harman, 1975; Wong, 2006) — is not what the round table stages, and is not what disables the mechanism. Harman's relativist bargains, persuades, and legislates; he titled the programme morality as politics.

The thesis concerns what makes a moral claim true rather than who has standing to press one; it is met where it belongs: at Chapter 14 (§6.4), where the anchor distinguishes real error from mere difference, and at Chapter 14 (§3.5), where it is placed on the partition. What the sections below diagnose is the practical stance the word has come to name in political life — the refusal of standing to persuade, on the ground that no one occupies the position from which to do it. The two are independent, since a Harman-style relativist may reject the stance, while the stance's adherents typically hold no metaethical view at all. It is the stance, not the thesis, that pairs with absolutism, because it is the stance that is authority-dependent.

All three positions share a structural defect: authority-dependence. Absolutism demands authority from above, since moral law is valid because its source is superior, which renders legislation unreformable. Relativism demands authority from below, since persuasion is illegitimate unless the speaker possesses special standing, which renders legislation impossible. The making of law requires opportunity rather than authority: the chance to earn assent through the merit of what is offered rather than the standing of who offers it.

The two laws of social equality are jointly violated, and in a determinate way. The first law is nominally upheld, since all are bound by the law, while the second law is specifically foreclosed: by absolutism because the law is unreformable, so no hierarchy can

reshape the standards, and by relativism because the law is unlegislatable, so no hierarchy is permitted to persuade another. In both cases the lawmaking mechanism of §5 is disabled, since the deliberative field in which private hierarchies would become shared standards is closed — by absolutism from above, by relativism from below. The second law, as §4.4 established, is the structural requirement that keeps the field open, and authority-dependence is what closes it.

Authority-dependence is the systematic prioritisation of derived value (the authority of the source) over primary value (the evaluative adequacy of the content), in the vocabulary of Chapter 15. The absolutist evaluates a moral claim by its derived value: does it come from God? From natural law? From the right tradition? The relativist evaluates a moral claim by its derived value: does the speaker have the right standing? Is she from the right culture? Does she have the right to judge? In both cases, the primary value — the evaluative adequacy of the claim's content — is subordinated to derived value.

7.2 What is wrong with relativism

Relativism may seem to promote equality and peace, while it paralyses the legislative process: by denying individuals the right to persuade one another, it suppresses moral debate and stalls the development of law. A relativist society upholds the first law and never attains the second, since no one can legislate, and therefore no one can progress.

The ideology rests on two mistaken assumptions. The first confuses social equality with equality of output. Ten bakers entering one competition stand as social equals, and their cakes still differ in taste, texture, and appearance; the equality of the makers settles nothing about the ranking of what they make. Moral claims, ideas, and arguments are outputs of the same kind, variable in quality however equal their authors. What social equality secures is an equal opportunity to compete — and the possibility that any participant's offering is the one the others come to recognise as best.

The second assumption holds that persuasion needs authority. The competition shows otherwise, since the judges assess each entry on its own merits, and a baker prevails through the quality of what she presents, with her standing contributing nothing. The challenge “Who are you to tell me what is right and wrong?” has the same shape as “Who are you to bake the best cake?”, and the same answer: no one in particular, and the offering may still persuade on its quality, refused only by a judge who declines to taste. The analogy carries the first assumption only, since cakes have judges and criteria and the relativist’s point is that morals have neither; against the second the answer is the anchor (Chapter 14, §6.4): the judge is the affected party, her registration the criterion, so the offering is assessed by its accuracy about her, whoever makes it; on the tracking reading her welfare is the criterion and her registration its report, so the answer stands. Relativism therefore silences moral debate in equality’s name, suppressing opportunity where only merit is needed.

7.3 What is wrong with absolutism

Absolutism renders law immutable, since nobody has the power to change it, no matter how necessary the change may be. It is a form of tyranny by proxy, where a “superior source” functions as a remote tyrant acting through its followers — Stage 4 of the anatomy of dogma (Chapter 17, §6) in political form: the standard severed from the foundation that gave it authority, self-authorising, a closed loop.

In many ways it can be more restrictive than tyranny itself. A tyrant may, at least in theory, be confronted or persuaded, whereas the author of an absolutist law is absent, inaccessible, or altogether non-existent, leaving no avenue for appeal, revision, or reform. In the terms of Chapter 17 (§7.1) the difference is the ratchet: a living tyrant can in principle still be persuaded, so the instrument that banks corrections still turns, whereas an absolutist code is the ratchet run in reverse, the layer that would hold corrections converted into the layer that holds them out.

The specific absolutist worldview — that all humans are equally inferior to a single superior authority — enshrines what may be termed equality by inferiority: the belief that no moral opinion, however

persuasive or well-reasoned, can ever match or surpass the decrees of a designated superior source. But where does this belief come from? For most absolutists the answer lies in the perceived origin of the law: it is either objective or infallible. Both claims collapse.

The objectivity claim confuses the form of the law with its authority. A written commandment is objective in form — visible, readable, repeatable — while objectivity in form does not imply moral authority, which is the same separation Chapter 16 (§2.1) drew between what fixes a standard's reference and what could grade anything by it. What gives a rule legal weight is not its clarity but consent: if a community agrees with its instruction and adopts it through persuasion, it becomes enforceable.

The infallibility claim was dismantled in Chapter 17, twice over. Evidentially, because to declare a law infallible is to assume its author is omniscient, and omniscience is impossible for any finite mind to verify (Chapter 17, §5.1). Structurally, because the omniscient guarantor requires the conjunction of the two infinities, defeated at theorem strength (Chapter 17, §5.2; Chapter 10, §3.2). The foundation is therefore unavailable rather than merely unproven.

7.4 The bipolar dynamic

These two ideologies present themselves as each other's antidote. Absolutism often emerges in response to relativistic lawlessness, portraying itself as a saviour ideology that restores moral order, while relativism arises in rejection of absolutist rigidity, seeking to escape dogma by denying the very notion of superior principles. Together they form what may be called bipolar authoritarianism. The structural claim is that the two poles share a single authority-dependent structure while repelling each other's content; that a society tends to swing between them over time is an empirical pattern the framework is consistent with and illustrates below, rather than a result it derives.

Yet they are not opposites. They are mirror images of a single authority-dependent structure. The absolutist says: "Without authority, there is no law." The relativist says: "Without authority, there should be

no law." Both agree that authority is the precondition for legislation; they disagree only about whether the precondition is met. Democracy rejects the precondition entirely.

The oscillation between the poles is an empirical pattern the framework illustrates rather than derives; it is widely observed. The Iranian experience illustrates it with particular clarity, since the Shah's secular authoritarianism provoked theocratic absolutism, each regime presenting itself as the antidote to the other's excesses. Post-Soviet Eastern Europe illustrates a different instantiation, since the collapse of communist absolutism produced moral fragmentation in some societies, creating a vacuum that nationalist populism filled by reimposing a different absolutism.

Contemporary Western democracies show the dynamic in real time, since liberalisation of social norms has been accompanied by backlash movements seeking to reimpose absolute standards. This is diagnosed here as a failure of authority-dependence rather than of liberalisation, since the backlash objects not to liberalisation's content but to the loss of the authority structure that previously constrained moral discourse. The political-science account of the same Western pattern — cultural backlash driven by status anxiety among declining majorities (Norris & Inglehart, 2019) — is complementary, since it names the sociological carrier of the swing while the diagnosis here names the structure the swing reaches for, and both readings mark the pattern as illustration rather than derivation.

8 Engagement with the Literature

8.1 Habermas and discourse ethics

The convergence with Habermas's (1990b) discourse ethics was introduced in the previous chapter, and the divergence is developed here. Habermas derives the conditions for legitimate discourse from the pragmatics of communication, while the framework derives them from the structural priority of deep persuasion, where the difference produces

a difference in what the conditions are conditions for: legitimate norms for Habermas, structural peace for the analysis.

The divergence matters because Habermas's account is compatible with permanent managed disagreement, since legitimate norms can be imposed by majority vote on a dissenting minority, and the imposition is legitimate if the deliberative process was fair. The framework identifies a further condition: preference-alignment. A norm imposed by majority vote after fair deliberation is a product of shallow persuasion rather than deep persuasion, so it may be legitimate without constituting peace. Habermas tells us when to stop deliberating and impose a norm; the framework tells us when to continue deliberating — until the preferences are aligned, or until alignment proves impossible and shallow persuasion is accepted as the best available outcome.

8.2 Rawls and overlapping consensus

Rawls's (1993) political liberalism seeks an overlapping consensus among holders of reasonable comprehensive doctrines. The present chapter is sympathetic to the Rawlsian aspiration. But it identifies a gap: Rawls presupposes that reasonable people will converge on principles of justice when they adopt the perspective of free and equal citizens. He does not specify the mechanism by which this convergence is achieved — deliberately so, because he wants his theory to be compatible with multiple accounts of convergence. The present account offers one such mechanism: the lawmaking mechanism of §5 — deep persuasion under the conditions of free speech, open speech, and democratic equality. The two laws of social equality provide the foundation for Rawls's "free and equal citizens": equal under the law (free from privilege) and equal above the law (free to shape the standards they live under). Rawls identifies what the consensus should contain; the analysis identifies conditions that make consensus-seeking possible.

8.3 Mouffe and agonistic pluralism

Mouffe's (2000, 2013) agonistic democracy was engaged in the previous chapter (§7.2), where her agonism divides along the two-case line: strategically accepted, it is sophisticated shallow persuasion; held

as fair, it is deep persuasion at the framework level, adversaries whose level-zero disagreement is housed inside a level-one alignment. What the account adds is the vocabulary to distinguish management from resolution: Mouffe correctly identifies what most democracies achieve, managed disagreement within a shared framework, a real political achievement; the diagnosis is that management is not peace, and that the conditions for peace are specifiable even if indefinitely far from achievement.

8.4 Institutional design: how the mechanism operates at scale

The three modes map onto institutional structures in a way Rawls (1993), Habermas (1996), and Sen (1999) have each partially developed. Rawls's basic structure is the institutional framework within which deep persuasion can operate, and the original position is a thought-experimental device for generating the engagement it requires, since imagining oneself behind the veil compels consideration of the evaluative experiences of those in very different social positions.

Habermas's (1996) Between Facts and Norms mirrors the lawmaking/enforcement distinction, since deliberation and coercion require different institutional structures and different normative justifications. Sen (1999) is the most concrete, since his capabilities approach identifies the material conditions — education, health, political participation — that must be in place before agents can engage in sustained evaluative deliberation: a population struggling with illness, illiteracy, or deprivation cannot conduct the exchange deep persuasion presupposes.

The mapping is: enforcement institutions (police, courts, military) address survival. Shallow persuasion institutions (constitutions, federal arrangements, incentive structures) address security. Deep persuasion institutions (deliberative assemblies, truth commissions, educational systems, free media) address alignment. A well-designed political system maintains all three but privileges the conditions for deep persuasion, because deep persuasion is the only mode that produces structural peace.

9 Objections and Replies

9.1 The cultural imperialism objection

Objection. The derivation of free speech, open speech, and democratic equality coincides with Western liberal democratic values. Is this cultural imperialism in philosophical disguise?

Reply. The derivation does not begin from liberal values and argue for them; it begins from an analysis of conflict resolution and derives conditions any society aiming at structural peace must satisfy. But the objection deserves more than this, and the scrutiny should be the opening move: if the framework yields Western liberal values, the question is whether it is structural, in which case the convergence is informative, or covertly parochial, in which case it is circular.

Benhabib (2004) treats the tension between universal norms and cultural particularity most rigorously, arguing that democratic iterations — the reinterpretation and contestation of universal norms within particular cultural contexts — are how universality and particularity are reconciled. On the present analysis her democratic iterations are a form of deep persuasion, and her account shows that the structural requirements do not dictate the institutional forms through which they are satisfied, since different cultures may satisfy the same requirements through different arrangements.

The strongest non-Western alternative must be engaged directly. Confucian social philosophy prioritises harmony (和) achieved through hierarchical order — filial piety, ritual propriety, deference to elders and authorities — rather than through the egalitarian exchange of evaluative perspectives the present derivation identifies. Its contemporary political form is Bell's (2015) meritocracy (§6.1). If Confucian harmony constitutes a structurally distinct path to peace that does not require free speech, open speech, or democratic equality, the derivation's universality claim fails.

The framework predicts that it does not. Harmony sustained by hierarchical authority is, on the three-mode taxonomy (Chapter 21), a sophisticated form of shallow persuasion: the emotional weight of filial

piety, social propriety, and respect for elders overrides conflicting preferences without transforming them. The conflicting preferences survive beneath the compliance. When the authority structure weakens — as it has repeatedly in Chinese history, through dynastic collapse, cultural revolution, and modernisation — the suppressed preferences resurface, producing the instability that shallow persuasion predicts. Deep persuasion would produce a harmony in which the preferences themselves are aligned, not merely overridden by the derived value of hierarchical authority.

A Confucian may object that filial piety transforms character, so the child who practises respect for her parents comes to feel the respect rather than merely perform it. The objection identifies a real phenomenon, and the framework accommodates it: moral education that cultivates empathic sensitivity (Chapter 16, §A.2) is a form of deep persuasion, and the Confucian emphasis on character formation through practice is structurally aligned with the analysis's own account of how evaluative hierarchies are shaped.

What the framework denies is that the hierarchical ordering of persons — the claim that some persons' evaluative perspectives are structurally entitled to override others' — is a condition for peace rather than an obstacle to it. Confucian harmony at its best is deep persuasion operating through moral education; Confucian hierarchy at its worst is shallow persuasion sustained by authority. The derivation targets the hierarchy, not the harmony.

A symmetric test. Stated so, the reply invites the charge the chapter levels elsewhere at its opponents: transformation is credited to deep persuasion and stability to shallow, so no Confucian datum could tell against the framework. The exposure is closed with a symmetric test.

The taxonomy's load-bearing variable is durability (Chapter 21): preferences transformed through engagement survive the removal of the sustaining structure, while preferences overridden by authority re-express when the authority weakens. The framework's classification of hierarchical harmony is therefore falsifiable by the Confucian case itself. A society whose harmony persisted undiminished through the

collapse or withdrawal of its authority structures — alignment surviving right where the shallow reading predicts resurfacing — would be evidence that hierarchical ordering transforms rather than overrides, and the classification, not the data, would have to yield. The historical record is offered as evidence under that test, and claims no immunity from it.

Two observations mitigate the circularity worry. First, Sen (1999) argues from within a non-Western intellectual context that freedom of expression and democratic participation are instrumentally and intrinsically valuable across cultures, so the convergence between the analysis's derivation and Sen's independently motivated argument, developed from Indian political philosophy, is evidence of general validity rather than cultural imperialism. Second, Appiah (2006) argues that cross-cultural moral conversation is possible without cultural imperialism, since the conversation itself is a form of deep persuasion, and its occurrence does not presuppose that one culture's conclusions are correct in advance.

The test is whether any step in the derivation smuggles in a culturally specific assumption. The derivation proceeds: peace requires deep persuasion; deep persuasion requires communication; communication requires free and open speech; free and open speech require equal standing. Each step is a claim about the requirements of a specific outcome, peace, so no cultural preference is being expressed. A non-Western society that achieves structural peace through a different mechanism would refute the derivation, which predicts that no such mechanism exists.

9.2 The incommensurability objection

Objection. Some moral disagreements involve incommensurable values, so deep persuasion may be impossible.

Reply. I. Berlin (1969) does not deny that moral claims can be compared; he denies that they can always be ranked on a single scale. This is value pluralism rather than relativism, and the distinction matters, since the relativist holds that no standpoint exists from which

one view can be judged better than another, while Berlin holds that comparison sometimes reveals real trade-offs rather than a single ranking.

The account is compatible with his pluralism, since the modality constitutes multiple oriented responses that may pull in incompatible directions, and the forced-choice diagnostic (Chapter 15, §4.1) reveals which prevails without pretending the losing value was illusory. Where incommensurability obtains at level zero, framework alignment one level up (Chapter 21, §2.1) remains the first resort; where that too is unavailable, shallow persuasion at its most sophisticated is the best remaining outcome. The diagnostic function stands, since the analysis identifies which conflicts are resolvable at some level and which require permanent management.

9.3 The stability objection

Objection. A society built on deep persuasion is inherently unstable because laws are subject to revision.

Reply. A society built on enforcement achieves stability of compliance without stability of preferences. A society built on deep persuasion achieves stability of preferences, which naturally produces stability of compliance. The latter is the more stable of the two. A law that cannot be revised is a law that cannot be corrected — and the inability to correct mistakes is a deeper form of instability than the ability to revise them.

10 Conclusion

The structural sequence of the book is complete. From a single observation — that content requires contrast — the analysis has established the self-vindication of difference and the closure of the grid, built on them the constructed accounts of value, morality, authorship, and purpose, and drawn the conditions for democratic peace from what deep persuasion is. Each link carries the grade of the weakest link beneath it.

The chapter yields three results, each graded as the Establishes line records and each inheriting the security of the empathic metaethics on which the whole political sequence rests (Chapters 14–16). What is necessary here is conditional necessity, binding on whoever seeks aligned social peace, and the Register's grades rather than a blanket modality carry the rest. The conditions for deep persuasion are established; the grounding of democratic equality is constructed, since it rests on the mechanism and the anchor; and the diagnosis of bipolar authoritarianism (§7) is constructed, inheriting the conflation, with the oscillation between the poles illustrated rather than derived.

The conditions for deep persuasion. The structural conditions — expressibility, receptivity, and non-exclusion — follow with necessity from what deep persuasion is. Free speech, open speech, and democratic equality are the institutions that secure them: the conditions are structural requirements rather than liberal values, with the institutions the best available means of meeting them.

The grounding of democratic equality. Democratic equality combines two gradings §4 keeps apart. Moral standing is grounded in the being's own actual or potential capacity for valenced states, with consideration graded by what is at stake for the being, which empathic intensity tracks. Equal deliberative standing is grounded in the mechanism, since a perspective either can or cannot enter the field, so every bearer of a hierarchy counts equally as a source of perspectives.

The two laws of social equality define democratic equality as the conjunction that preserves the conditions for the mechanism, where the second law's derivation from the mechanism itself — rather than from autonomy, fairness, or relational respect — yields the diagnostic consequence behind the bipolar analysis: both relativism and absolutism disable it through the same defect of authority-dependence.

Democracy as structural form. Democracy is the political structure that preserves all three conditions, and democratic failures are diagnosed as failures of specific conditions. The framework converges with Habermas on the importance of fair deliberation, with Rawls on the aspiration to overlapping consensus, and with Mouffe on the realism of

managed disagreement — while providing the vocabulary to distinguish management from resolution.

The remaining chapters apply these results to religion, the most socially consequential form of absolutism in contemporary societies. Democratic equality is not only essential for open-minded debate and the legislative process; it also shapes how we judge and treat one another in everyday life. An authority-based ideology does not merely distort the legal system, since it corrupts human relationships. And the least tractable form of corruption operates behind a wall of private intolerance that only speech can dismantle, rather than in the public square where law can reach. The next chapter identifies how religious absolutism constructs that wall, and what bringing it down would require.

Grades and Dependencies

1. Modal weights. The three conditions are *established*; standing and the two-cell diagnosis are *constructed*, by inheritance; the lawmaking mechanism is *constructed*; its democratic identification — that democracy is the structure that houses it — is *proposed*, since other arrangements could in principle realise the same operation (§5.2). All necessity here is conditional: binding on whoever seeks aligned social peace (Ch. 21, §1), and the Register marks the three rows with that condition (note s), since the route to them runs through Chapter 21's constructed modes. The institutional layer is best-available means, not entailment (§1). The pole-to-pole oscillation is illustrated, not derived (§7.4). The Premise Register carries the chain.

2. Scope and exposure. The standing criterion fixes whose states count, not how conflicting standings weigh — the abortion debate's substantive work lies in the weighing, and the dispute with McMahan belongs at the metaphysics of the continuant (§4.1). Boundary cases — foetal, AI, vegetative — are assigned to evidence (§4.2). Animals hold full graded consideration plus a claim to faithful representation; deliberative equality holds among contributors (§4.3). The falsifiability concession of §A.3 is real, and the Confucian classification submits to the durability test it applies to others (§9.1).

Appendix: Applications

This appendix holds the applications of the chapter's results: the democratic failures read as failures of the three conditions (§A.1), the institutional implications the conditions carry, with the predictions the Register stakes on them (§A.2), and the objection aimed at the failures (§A.3). No later result rests on it.

A.1 Democratic failures

Democracies have produced colonialism, populist authoritarianism, tyranny of the majority, and the suppression of minorities. The analysis's response is structural, since each failure is a failure of one or more of the three conditions.

Colonialism denied the second law entirely, since the colonised had no opportunity to bring their hierarchies into the deliberative field: the colonisers equal above the law among themselves, the colonised excluded from the field altogether. Populist authoritarianism collapses the second law to one person, since dissent is treated as a threat to be suppressed rather than a secondary object to be registered, and the leader's hierarchy is the only one that shapes the standards. Tyranny of the majority violates it in the second direction, since the majority's preferences are imposed through the sheer weight of numbers rather than through deep persuasion. In each case the first law may be nominally preserved while the second is violated, and with it the lawmaking mechanism is disabled in the relevant domain.

A.2 Institutional implications

The structural conditions derived in §§2–5 generate specific predictions about what kinds of institutions preserve or disable the conditions for deep persuasion. These predictions are not a political programme but a set of diagnostics — tools for identifying when an institution's design is compatible with the conditions for democratic peace and when it is not.

An institution that evaluates persons by creed rather than conduct disables open speech, since standing conditioned on doctrinal allegiance

rather than the quality of evaluative engagement excludes the very perspectives most in need of being heard: dissenters, minorities, outsiders. The prediction is specific — institutions that condition participation on creed will exhibit lower rates of preference-transformation and higher rates of enforced or shallow compliance.

An institution that suppresses minority perspectives disables free speech. The structural prediction: institutions with stronger protections for minority speech will exhibit more instances of preference-transformation over time, because the perspectives most likely to transform majority preferences are those that differ most from them.

An institution that treats its own standards as infallible disables the revisability deep persuasion requires. The prediction: institutions that build revisability into their foundational commitments (constitutional amendment procedures, periodic review, sunset clauses) will exhibit greater capacity for deep persuasion than institutions that declare their foundations unreformable.

An institution that separates lawmaking from law enforcement preserves the distinction between the empathic foundation and the orthotic superstructure. Where one body both creates and enforces the rules, the orthotic operation tends to displace the empathic one, because enforcement is more efficient and less demanding than deliberation. The prediction: institutions with firm separation between legislative and enforcement functions will exhibit higher rates of empathic engagement in the legislative process.

These predictions are testable in principle against comparative institutional data and are not yet operationalised: "rate of preference-transformation" and "empathic engagement in the legislative process" have no shared measures, and until they do the predictions are directions for research. The framework identifies the structural conditions from which they follow and offers them as hypotheses for institutional analysis.

A.3 The democratic failure objection

The objection is that every failure of democracy is attributed to insufficient democracy, so that the claim is unfalsifiable.

The objection has force, and candour requires acknowledging its practical weight. In principle the claim is falsifiable, since it would be refuted by a case in which all three conditions are fully satisfied and the lawmaking mechanism nonetheless fails to produce any movement toward preference-alignment. In practice the three conditions are never fully satisfied in any existing society, which means the falsification condition is extremely difficult to meet.

This is a real limitation of the structural analysis's testability in political contexts, analogous to the difficulty of testing ideal-gas laws when no perfect gas exists. Its value is diagnostic rather than predictive, since it identifies what the conditions for peace are and where existing societies fall short rather than predicting that any society will achieve peace. The diagnostic yield is what makes the framework useful, not a guarantee that the ideal is achievable.

References

Anderson, E. (1999). What is the point of equality? *Ethics*, 109(2), 287–337.

Appiah, K. A. (2006). *Cosmopolitanism: Ethics in a world of strangers*. Norton.

Bell, D. A. (2015). *The China model: Political meritocracy and the limits of democracy*. Princeton University Press.

Benhabib, S. (2004). *The rights of others: Aliens, residents, and citizens*. Cambridge University Press.

Berlin, I. (1969). *Four essays on liberty*. Oxford University Press.

Bloom, P. (2016). *Against empathy: The case for rational compassion*. Ecco.

Brennan, J. (2016). *Against democracy*. Princeton University Press.

Christiano, T. (2008). *The constitution of equality: Democratic authority and its limits*. Oxford University Press.

Cohen, J. (1989). Deliberation and democratic legitimacy. In A. Hamlin & P. Pettit (Eds.), *The good polity: Normative analysis of the state* (pp. 17–34). Basil Blackwell.

Donaldson, S., & Kymlicka, W. (2011). *Zoopolis: A political theory of animal rights*. Oxford University Press.

Dotson, K. (2014). Conceptualizing epistemic oppression. *Social Epistemology*, 28(2), 115–138.
Dworkin, R. (1977). *Taking rights seriously*. Harvard University Press.
Estlund, D. M. (2008). *Democratic authority: A philosophical framework*. Princeton University Press.
Fishkin, J. S. (2009). *When the people speak: Deliberative democracy and public consultation*. Oxford University Press.
Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.
Gutmann, A., & Thompson, D. (2004). *Why deliberative democracy?* Princeton University Press.
Habermas, J. (1990b). *Moral consciousness and communicative action*. MIT Press.
Habermas, J. (1996). *Between facts and norms*. MIT Press.
Harman, G. (1975). Moral relativism defended. *The Philosophical Review*, 84(1), 3–22.
Holmes, O. W. (1919). Dissent in *Abrams v. United States*, 250 U.S. 616.
Korsgaard, C. M. (2018). *Fellow creatures: Our obligations to the other animals*. Oxford University Press.
Landemore, H. (2013). *Democratic reason: Politics, collective intelligence, and the rule of the many*. Princeton University Press.
Langton, R. (1993). Speech acts and unspeakable acts. *Philosophy & Public Affairs*, 22(4), 293–330.
Maitra, I., & McGowan, M. K. (Eds.). (2012). *Speech and harm: Controversies over free speech*. Oxford University Press.
Marquis, D. (1989). Why abortion is immoral. *Journal of Philosophy*, 86(4), 183–202.
McMahan, J. (2002). *The ethics of killing: Problems at the margins of life*. Oxford University Press.
Meiklejohn, A. (1948). *Free speech and its relation to self-government*. Harper & Brothers.
Mill, J. S. (1859/1999). *On liberty* (E. Alexander, Ed.). Broadview Press.
Mouffe, C. (2000). *The democratic paradox*. Verso.
Mouffe, C. (2013). *Agonistics: Thinking the world politically*. Verso.
Norris, P., & Inglehart, R. (2019). *Cultural backlash: Trump, Brexit, and authoritarian populism*. Cambridge University Press.
Nussbaum, M. C. (2006). *Frontiers of justice: Disability, nationality, species membership*. Harvard University Press.
Pettit, P. (1997). *Republicanism: A theory of freedom and government*. Clarendon Press.
Rawls, J. (1993). *Political liberalism*. Columbia University Press.
Regan, T. (1983). *The case for animal rights*. University of California Press.

Schauer, F. (1982). *Free speech: A philosophical enquiry*. Cambridge University Press.
Scheffler, S. (2003). What is egalitarianism? *Philosophy & Public Affairs*, 31(1), 5–39.
Sen, A. (1999). *Development as freedom*. Knopf.
Singer, P. (1975). *Animal liberation*. New York Review/Random House.
Thomson, J. J. (1971). A defense of abortion. *Philosophy & Public Affairs*, *1*(1), 47–66.
Waldron, J. (1999). *Law and disagreement*. Clarendon Press.
Waldron, J. (2006). The core of the case against judicial review. *Yale Law Journal*, 115(6), 1346–1406.
Wong, D. B. (2006). *Natural moralities: A defense of pluralistic relativism*. Oxford University Press.
Young, I. M. (2000). *Inclusion and democracy*. Oxford University Press.

CHAPTER 23

The Democratic Deficit of Theocracy

Presupposes: the ground/mind conflation (Ch. 10); the great moral conflation (Ch. 16); the perfection/infallibility distinction (Ch. 17); bipolar authoritarianism (Ch. 22).

Establishes: the divine wall of discrimination; the three structural transformations required for democratic compatibility; what remains for religion after them.[1]

In brief. Laws can ban segregation and police the workplace; they cannot compel friendship, love, or marriage. Behind that boundary — the wall of private intolerance — the only tool with power is speech; the form most resistant to it is the belief that someone is not good enough to be *right*. The divine wall is that belief given eternal sanction: God separating souls by creed and not character, the celestial line projected onto earth. Conditional on the empathic constitution of morality (Ch. 16), the argument derives three transformations — theocratic to humanistic, absolutist to democratic, creed-based to merit-based — already under way inside the traditions. What is removed is the claim to absolute authority, not faith.

---

1 Introduction

This chapter applies the analysis to the most socially consequential form of absolutism: religious absolutism. It argues that theocratic morality generates a specific structural deficit in democratic capacity through a mechanism the analysis is uniquely positioned to identify: the elevation of creed-based evaluation over character-based evaluation, sanctioned by the doctrine of divine judgement by allegiance.

The argument develops the concept of the divine wall of discrimination: the structural consequence of a doctrine in which God separates souls by creed rather than character, projected from celestial judgement onto earthly social relations (§3). It first identifies four convergent results from earlier chapters that bear on theocratic morality (§2), then locates the divine wall within the broader spectrum of discrimination and the progression from the jungle state through the zoo state toward the home state (§§2.1–2.2). From there it derives three structural transformations required for religious compatibility with democratic peace (§4), and examines what remains for religion after these transformations (§5).

A note on tone. This chapter engages with doctrines that are held with deep sincerity by billions of people. The chapter's argument is structural: it targets specific doctrinal features (creed-based judgement, scriptural infallibility, the ontological superiority of God over humanity) and derives consequences from them. It does not target faith as such, devotion as such, or religious community as such. The chapter's obligation is to reconstruct the strongest self-understanding of the traditions under critique before identifying the consequences the analysis reveals. Where the argument falls short of this obligation, the failure is the chapter's, not the traditions'.

A note on conditionality. The argument that follows is conditional. If morality is constitutively empathic, as Chapter 16 argued, then creed-based evaluation is a conflation of orthotic with empathic evaluation, and the consequences identified in this chapter follow. A theist who rejects the empathic constitution of morality, holding instead that morality is constituted by divine command, can resist the chapter's conclusions by resisting the premise. The chapter's target is the credal strand within the traditions; a conduct-centred tradition (§3.1) has little to revise.

The conditional structure is maintained throughout, since each key derivation proceeds from the premise "if morality is constitutively empathic, then…". The chapter does not pretend to neutral ground, because it develops consequences of the established results. A theist who finds the premise implausible should therefore direct her critique to Chapter 16, where the premise is argued for on four independent structural grounds. The chapter's obligation is to show that if the premise holds these consequences follow, and that the consequences are significant enough to warrant the theist's sustained engagement with that premise.

2 Four Convergent Results

The preceding chapters have established four results that converge on theocratic morality.

First, the incoherence of origination from non-being. The grounding terminus is unintelligent; creation pictured as a transition from non-being is structurally incoherent, while creation as sustaining dependence survives (Chapter 9, §2.3); divine simplicity cannot be conjoined with intelligence. Second, the constitutively empathic character of moral evaluation. Morality is empathic response to the emotional consequences of conduct for others; orthotic assessment is derivative (Chapter 16). Third, bounded authorship. The unified causal model shows that no conscious agent authored her evaluative hierarchy; the perspectival asymmetry shows that absolute moral responsibility is structurally incoherent (Chapter 18). Fourth, the authority-dependence of absolutism. Absolutism disables the conditions for deep persuasion by foreclosing the legislative process — violating the second law of social equality, the requirement that keeps the deliberative field open to every perspective (Chapter 22, §4.4).

The convergence generates consequences that no individual result produces alone. Consider a religion that claims a God who produced being from its absence, grounds morality in divine command, presupposes unbounded authorship to justify eternal punishment, and treats its doctrines as unreformable. Each of its central claims is independently undermined, and because the four combine, the effect on democratic capacity is structurally significant.

2.1 The spectrum of discrimination and the wall of private intolerance

Before developing the divine wall, the chapter must locate it within a broader structural analysis of discrimination. Discrimination is the act of judging someone as inadequate or unworthy based on a partial or incomplete evaluation, typically reducing a person to a single feature while ignoring the rest. Intolerance is a form of discrimination, expressed across a broad spectrum. At the public end, people may be killed, imprisoned, or otherwise silenced for who they are. A society can enact laws to restrict or punish these public forms: by banning racial segregation, protecting workplace rights, or enforcing equal access to services.

But private intolerance is harder to reach, because laws cannot compel friendship, love, or consensual marriage with those against whom we harbour prejudice. This legal boundary, the wall of private intolerance, cannot be dismantled by force, so behind it the only tool with real power is speech.

Yet even persuasion becomes difficult in the face of the form of discrimination most resistant to it, found at the most private extreme of the spectrum: the belief that someone is not good enough to be right. This is the core of closed-mindedness — to reject a person's claim because of who is making it rather than because of its merit — and it is what divides a society most, since it is refusal rather than disagreement. That refusal echoes in both relativism and absolutism, two ideologies that, in different ways, disqualify the speaker before the idea is heard.

The divine wall, as §3 will show, operates at this most private end of the discrimination spectrum, where law cannot reach and only deep persuasion has power.

2.2 Three states of social integration

The progression from intolerance to peace can be understood through three states. The first is the jungle state: open conflict, violent imposition of creed, beatings, killings, imprisonment. Communities fight to impose their moral frameworks on one another.

Eventually, exhausted by violence, the communities agree to separate: building walls, drawing borders, establishing rules of non-interference. This is the zoo state, where animals coexist harmoniously so long as they are kept in cages. Legal enforcement holds public discrimination at bay while walls of private intolerance continue to divide communities across countries, regions, cities, and boroughs. Many would claim this is an acceptable endpoint. By the analysis of Chapter 21, however, a walled coexistence is managed conflict, because the contained preferences persist and the arrangement holds only while its overriders hold.

The zoo state should be treated as temporary, because it is possible to move on toward a time when people interact as they do within their

own communities. Not merely in a state of law-enforced public coexistence, but in one of widespread private acceptance. This is the home state, where individuals treat one another with respect, care, and an absence of prejudice, as fellow members of a single community rather than as strangers held apart.

The contact evidence. The zoo-to-home progression is the contact hypothesis in structural dress, and the social-psychological literature that has tested it should be named, since it supplies the empirical spine the imagery lacks. Allport (1954) proposed that contact reduces prejudice under specifiable conditions — equal status between the parties, common goals, cooperation rather than competition, and the sanction of authorities and custom — and half a century of accumulated studies confirms the effect meta-analytically, with the largest reductions where the contact is personal, sustained, and capable of producing friendship (Pettigrew & Tropp, 2006).

The base carries a measured qualification, because confined to experimental designs the effects run smaller, with delayed outcomes rarely measured (Paluck, Green, & Green, 2019). The spine therefore supports the progression at the strength the trials license and no more.

Allport's conditions read, in the present vocabulary, as the operating requirements of deep persuasion across a wall. Equal status is standing (Chapter 22, §4). Cooperation toward common goals is the shared field in which hierarchies are exchanged. Institutional sanction is the zoo's legal scaffolding doing its one legitimate job: holding the space open while the walls come down from inside.

What the framework adds to the hypothesis is the mechanism underneath the correlation. Contact works when and because it converts the other from a category into a registered evaluative field, which is why superficial or competitive contact fails where Allport's conditions predict failure, and why the home state cannot be legislated into being but only made structurally available.

The progression from zoo to home requires that dialogue resume, since if it does not, the zoo remains the only alternative to the jungle. To move toward lasting peace, where conflict is neither merely suspended

nor sealed off, people must speak freely, listen openly, and confront opposing views face to face. It is only through deep persuasion that the walls can fall. The divine wall is the most structurally entrenched of these walls, because it is sanctioned by the doctrine of divine judgement itself rather than by human authority.

3 The Divine Wall of Discrimination

The chapter's central argument can be stated formally before the detailed analysis that follows:

(W1) Morality is constitutively empathic: what makes an evaluation moral is the evaluative modality's oriented response to the emotional consequences of chosen conduct for others. [Argued in Chapter 16, where it carries as constructed.]

(W2) Creed-based evaluation is orthotic: it checks persons against a doctrinal standard (correct belief, correct allegiance) and registers conformity or non-conformity.

(W3) Orthotic assessment is not moral evaluation. [From W1: moral evaluation requires empathic engagement with the emotional consequences for others, whereas orthotic assessment requires only conformity-checking.]

(W4) Therefore creed-based moral evaluation, the evaluation of persons' moral worth by their doctrinal allegiance, is the great moral conflation in its most consequential social form.

The remainder of §3 develops the social consequences of this conflation.

3.1 The structure of creed-based judgement

In many religious traditions, God is believed to judge human beings ultimately by their allegiance to Him. Those who believe in God and follow His commandments are destined for salvation, while those who wilfully and conclusively reject Him face eternal separation or punishment. The specific criteria vary across traditions, since some emphasise belief, others practice, others a combination, though the

structure under examination is constant: when the primary criterion of divine evaluation is creed, what you believe about God, rather than character, how you treat other persons, the consequences described below follow.

A preliminary qualification is essential for the chapter's structural integrity. Traditions vary enormously in the weight they assign to creed versus character. At one end of the spectrum, some traditions are primarily orthopractic: they evaluate persons mainly by conduct, with doctrinal belief playing a secondary role. Judaism's emphasis on mitzvot (commanded actions) and Buddhism's emphasis on the Eightfold Path place conduct at the centre of moral evaluation. At the other end, some traditions are primarily credal: much of evangelical Protestantism's sola fide (faith alone) doctrine, and certain strands of Sunni and Shia Islam that condition salvation on shahada (declaration of faith), place creed at the centre. Most traditions occupy intermediate positions. The chapter's argument applies most directly to traditions in which creed-based evaluation is primary, and applies with diminishing force as the tradition moves toward the conduct-based pole. A tradition that already evaluates persons primarily by character and not creed has less to revise.

Creed-based evaluation is orthotic assessment applied to persons. The person is checked against a doctrinal standard — right belief, right commands, right community — with conformity registered as moral worthiness and non-conformity as failure, while the empathic evaluation of how she treats others is structurally secondary.

Plantinga (2000) develops the most sophisticated epistemology in the neighbourhood. On his account the sensus divinitatis produces warranted belief in God; its failure to operate is a cognitive malfunction, much as colour-blindness is a perceptual deficit. That is a claim about warrant alone, and Plantinga nowhere infers the second from the first; reformed epistemology, read on its own terms, builds no wall.

The analysis engages the position only to mark the inference the divine wall requires and the epistemology does not license: from "her faculty malfunctions" to "her moral standing is diminished." Even if the

sensus divinitatis is a cognitive faculty, the question is whether its deliverances should take priority over empathic evaluation in the assessment of persons. A person with perfect colour vision and no compassion is not thereby morally superior to a colour-blind person of profound empathic sensitivity, and a cognitive achievement, however real, is not the criterion the empathic constitution of morality makes primary. The wall is built where a tradition makes that inference; the analysis's answer, grounded in Chapter 16, is that the inference is the conflation itself.

Plantinga's (1995) defence of exclusivism draws the same line from the other side. He argues that holding one's religious beliefs true, and the incompatible beliefs of others false, is neither arrogant nor unjustified even in full awareness of the piety those others display, and the chapter grants the point entire, since exclusivism about truth is an epistemic stance and no wall. The wall is exclusivism about standing: the further step from "her creed is false" to "her moral worth is less", a step Plantinga's defence does not take. A tradition can hold its doctrine true and its rivals mistaken while evaluating persons by character, which is what the third transformation (§4.3) asks of it and all it asks.

A stronger opposition comes from Hauerwas (1981), who argues that the church's distinctiveness from secular culture is itself a moral achievement: the community's moral life is constituted by its particular narrative, so dissolving creed-based boundaries would destroy the community's capacity for moral formation.

The response must distinguish two operations the Hauerwasian position runs together. The first is creed-based community formation, the voluntary association of persons around shared commitments. The second is creed-based moral evaluation of persons outside the community, the judgement that a non-member's moral standing is diminished by her non-membership. The first is legitimate, since voluntary association around shared evaluative orientations is itself a form of deep persuasion, and the chapter does not target it. The second is what the divine wall identifies.

Hauerwas's strongest case supports the first without entailing the second, because a community that forms around shared commitments can recognise the moral standing of those outside it without compromising its own identity. A believer who treats others badly may be forgiven if her creed is correct; a non-believer who treats others well may be condemned if her creed is wrong. This is the great moral conflation in its most consequential social form.

Creed as proxy for character. A further defence of creed-based evaluation must be addressed, because it is the strongest available. The theologian may argue that creed matters less as a direct moral criterion than as a predictor of character: what you believe about God shapes how you treat people, so evaluating by creed is a practical proxy for evaluating by character.

The response has two steps. First, the correlation between creed and character is empirical and defeasible, since non-believers can have excellent character, believers can have terrible character, while the correlation varies enormously across individuals, communities, and historical periods. Creed is therefore an unreliable proxy for character.

Second, and more fundamentally, relying on a proxy when the thing itself is directly available is the great moral conflation in its predictive form. Character — how a person actually treats others, whether she alleviates suffering, whether she promotes flourishing — is directly observable. Creed, what a person believes about God, is a secondary object whose derived value, its association with tradition, authority, and community, may or may not predict the primary value, the person's actual empathic conduct. Evaluating by creed when character is available resembles evaluating a gymnast's performance by the reputation of her coach instead of by her score: the proxy may carry information, while the direct measure is both available and more reliable.

Creed-based judgement systematically prioritises derived value, the value a person acquires through association with the right doctrinal community, over primary value, the value constituted by her actual conduct toward others. The divine wall is the structural expression of

this inversion: derived value overrides primary value in the evaluation of persons.

3.2 From celestial judgement to earthly division

When God is believed to separate souls into eternal factions according to allegiance, religious communities inherit this celestial line of separation and project it onto earthly relationships. The doctrine does not merely describe a post-mortem event, because it constitutes a present-tense evaluative framework. If God judges persons primarily by creed, then creed becomes the primary criterion by which the faithful evaluate one another in this life. The result is a doctrinal divide between the "blessed" and the "doomed", a partition more profound than any based on character, conduct, or contribution, because it carries the weight of eternal consequence.

This is the divine wall of discrimination. The wall operates through evaluative separation rather than hatred: a structural boundary between those who are on the right side of the divine standard and those who are not. It operates at the most private level of intolerance, since what it holds is not merely that someone is bad but that someone is not good enough to be right — that a non-believer's moral perspective is disqualified by her creed, regardless of its content.

In the vocabulary of the preceding chapter this is a violation of non-exclusion. Equal deliberative standing turns on the fact that a perspective's eligibility to enter the field as a candidate for uptake does not depend on prior approval of its source (Chapter 22, §4), and the wall withholds that very eligibility. The divine wall is therefore not only the great moral conflation in social form but the religious instance of the exclusion that disables receptivity, foreclosing deep persuasion across the doctrinal line at its root.

The witnessing defence. The strongest defence of creed-based evaluation must be addressed here, because it is the response that intelligent, compassionate believers are most likely to offer, and because it accepts the chapter's own terms.

The defence runs as follows. “You have misidentified our motivation. We do not evaluate non-believers by creed because we are performing an orthotic check. We evaluate them by creed because we empathically care about their eternal fate. If I really believe my neighbour will suffer eternally for her unbelief, then my urgency about her creed is empathic concern — I am responding to the emotional consequences of her situation for her. Witnessing, evangelism, and even social pressure to convert are expressions of empathic love rather than orthotic conformity-checking. A doctor who insists you take medicine you do not want is not being orthotic; she is being empathic. We are the doctor.”

This is the strongest version of the creed-based defence, because it does not retreat to “God said so”, which the great moral conflation analysis has already addressed. It claims instead that creed-based urgency is empathic urgency, real concern for the other’s suffering, and therefore satisfies the chapter’s own criterion for moral evaluation. The defence deserves three responses, each operating at a different level.

First, the empathic element must be acknowledged without reservation. A believer who fears for her neighbour’s eternal fate is experiencing a real emotional response: a real oriented attraction toward the neighbour’s well-being and a real repulsion at the prospect of the neighbour’s suffering. The evaluative modality is engaged. This is not pseudo-empathy, directed at an object that does not feel; it is empathic concern for a person the believer loves. The chapter does not question the sincerity or the emotional reality of the response.

Second, the structural problem must be identified. The empathic response is directed at a consequence, eternal damnation, whose reality depends entirely on a doctrine, creed-based divine judgement, that is itself the product of the orthotic layer. The believer empathically fears for her neighbour’s fate, while the fate she fears is constituted by the very doctrinal framework the chapter is examining. The empathy is real; its object is doctrinally constituted.

The analogy can be run the other way. A doctor who insists you take medicine for a disease she diagnosed is motivated by empathic

concern for your suffering. But if the disease exists only within the doctor's diagnostic framework, which is what is being questioned, then the empathic urgency does not validate the framework, because it is a product of it. The concern is real; it does not establish that the condition it responds to is.

Third, the circularity must be made explicit. The doctrine of creed-based judgement creates the danger, eternal separation for unbelief. The empathic response to the danger is then cited as justification for the doctrine's social consequences: pressure to convert, creed-based evaluation of persons, the divine wall. So the circle runs: the doctrine posits a suffering that the believer empathically responds to, and the empathic response is used to justify the social structure the doctrine generates. But the suffering is internal to the doctrinal framework, since a non-believer who does not accept the doctrine does not face the danger the believer fears for her. The empathic urgency is therefore urgency about a consequence the believer's own framework has posited, directed at a person who does not share the framework.

The structural parallel is complete. It is as though a doctor diagnosed a disease that only her diagnostic manual recognises, prescribed treatment the patient does not want, and cited her concern for the patient's health as justification, while the patient, examined by every other diagnostic framework, is perfectly well. The empathic element is real in the doctor; the disease is an artefact of the framework.

The witnessing-as-empathy defence therefore does not escape the divine wall analysis. It reveals the wall's most intimate mechanism, since the doctrine generates a fear, the fear generates empathic urgency, so the urgency sustains the social structure — all within a framework the non-believer has no reason to accept. The believer's love is real; the condition it responds to is doctrinally constituted; and the social consequence, the evaluation of the non-believer's moral standing by her creed, is the divine wall in its most compassionate and therefore most resistant form.

One limit of this diagnosis must be owned where it operates, rather than only in the chapter's preamble. Against the believer inside the

framework the circularity argument does not persuade, because for her the danger is real, the "doctrinally constituted" description is the disagreement restated, and the doctor analogy begs what she disputes — her manual, she holds, is the one that diagnoses truly.

The argument's addressee is therefore the reader assessing the framework and the doctrine from outside, and its force is the conditional force announced at the outset: if morality is constitutively empathic (Chapter 16), the witnessing defence fails to convert creed-based evaluation into empathic evaluation. Deep persuasion of the believer herself must travel through that premise and her engagement with it, rather than through this section's diagnosis, which maps the structure of her position without supplying her a reason she recognises to leave it.

3.3 Conditional divine love

The divine wall generates a further consequence: the conditionality of divine love. Many traditions claim that God's love is unconditional. But if judgement awaits those who make the "wrong" choice, and if hell is the consequence of sincere disbelief, then the love is conditional on allegiance, and the structural reframing holds regardless of the language in which the doctrine is expressed.

Stump (2010) argues that God's love is unconditional on creedal allegiance, directed toward the beloved's flourishing — an account structurally aligned with the empathic constitution of morality, since God loves persons because of who they are rather than what they believe. If Stump is right, the conditional-love consequence is not a feature of the deepest theological understanding but a distortion introduced by the creed-based overlay. The convergence is structural while the divergence is metaphysical, because Stump presupposes a creator God and the analysis does not.

The analogy with parental love makes the consequence vivid. Consider a parent who says: "When you turn thirty, I will judge you. If you have not done what I expected, I will abandon you — forever." No one would call this unconditional love. The structure is identical whether the judge is a parent or a deity.

Hick's (1989) pluralist hypothesis is the most developed philosophical alternative to creed-based exclusivism: the major traditions are different culturally conditioned responses to the same transcendent reality, "the Real", so salvation is available through all of them. Hick dissolves the divine wall by removing the creed-based criterion, arriving at the right structural conclusion through a different route — the phenomenology of religious experience rather than the orthotic/empathic distinction.

The divergence is equally important. Hick posits a transcendent Real behind the traditions, a postulate the analysis does not require, since its convergence on merit-based evaluation follows from the empathic constitution of morality and does not depend on whether a Real exists. A theist who accepts Hick's pluralism and one who accepts the framework arrive at the same practical conclusion for different reasons; the practical convergence matters more than the metaphysical disagreement.

The conditionality of divine love reinforces the divine wall by providing its ultimate sanction. To love someone across the wall (a non-believer, a member of another faith) is to love someone whom God has, or will, condemn. Such love is not merely difficult; it exists in structural tension with the doctrine of creed-based judgement. The divine wall does not prohibit love across its boundary; it makes such love structurally costly, requiring the believer to hold in tension her love for the person and her acceptance of the person's condemnation.

3.4 The interfaith test case

The most vivid manifestation of the divine wall is the struggle surrounding interfaith marriages. Such unions could serve as among the greatest catalysts for social integration and peace: when people of different faiths fall in love and build families together, the boundaries between communities dissolve organically. Repeated across generations, such unions erode creed-based division at its base.

But the divine wall makes interfaith marriage structurally difficult. The believing community must reconcile the member's love for a

person with that person's position on the wrong side of the doctrinal divide. The result is predictable: pressure to convert, conditions on the marriage, social distance from the non-believing partner's family, or outright refusal.

The social consequences are empirically observable. Putnam and Campbell's *American Grace* (2010), the most comprehensive empirical study of religion and social life in the United States, found that personal acquaintance with members of other faiths corresponds to warmer attitudes toward those faiths. It found interfaith marriage the most intensive form of such acquaintance. The Pew Research Center's Religious Landscape Studies corroborate the finding at the population level (Pew Research Center, 2015).

These are correlations rather than demonstrations of causation, which the chapter presents them as. But the correlations are consistent with a structural prediction: interfaith marriage dissolves the divine wall by creating families in which creed-based evaluation is structurally impossible, so the dissolution reduces the private intolerance the wall maintains.

A complication internal to the same study must be registered, because it appears to cut the other way. Putnam and Campbell's larger finding is that religion in the contemporary United States both divides and bridges: dense networks of religiously diverse personal ties, their "Aunt Susan" and "My Friend Al" effects, generate unusually high interfaith tolerance. Religiously observant Americans are, they find, on average more accepting of other faiths rather than less.

Taken flat, this looks like evidence against the divine wall. But the two findings operate at different levels, and the framework predicts their co-occurrence rather than being embarrassed by it. The wall is a property of creed-based doctrine, the evaluation of persons by allegiance, rather than of religious sociality as such, while bridging is produced by the mechanism this chapter endorses: personal contact across the doctrinal line.

What Putnam and Campbell record is therefore the wall being dissolved from below by interfaith acquaintance faster than doctrine

rebuilds it — deep persuasion operating through the very contact that such ties supply. Read as a claim about religious people's attitudes, the bridging data would tell against the thesis. Read as a claim about the structural tendency of creed-based evaluation, the bridging effect is the wall coming down rather than evidence that it was never standing.

A test the thesis can fail. Stated so, the reading courts the structure it must not have, where tolerance counts as the wall coming down and intolerance as the wall standing, so that every observation confirms. The exposure therefore calls for a test the thesis can fail.

The divine-wall thesis is a claim about doctrine, so it predicts a doctrinal gradient. Private intolerance, and resistance to interfaith intimacy above all, should track the creedal weighting of the tradition — highest where standing before God is conditioned on allegiance, lowest in the orthopractic and universalist strands. It should persist among the doctrinally committed even at high levels of interfaith acquaintance. A bridging pattern flat across that gradient, with exclusivist and orthopractic traditions equally warmed by equal acquaintance and doctrine adding nothing once contact is controlled, would be evidence that the wall is an artefact of contact scarcity rather than a doctrinal structure. The thesis would then be wrong, with reinterpretation unavailable. Putnam and Campbell's own comparisons across traditions are consistent with the gradient, though their design does not isolate it. The point here is the test's shape rather than its verdict: the thesis stakes itself on doctrine's residual effect after contact, and can lose there.

The survival objection. The case as so far stated is one-sided, and the objection it invites should be met on the analysis's own terms rather than waved past, because at first sight it turns the framework's central commitment against the framework's conclusion.

A member of a small or historically threatened tradition may resist interfaith marriage not out of creed-based contempt for the other but out of concern for the survival of a community whose continuity depends on transmission. The dissolution this section celebrates is, from within that concern, the slow disappearance of a people, a language of worship,

a form of life. This is not obviously the divine wall. It presents itself as an empathic consideration in its own right: a response to the anticipated suffering of loss, of ancestors' sacrifice rendered vain, of a grandchild who will not know the tradition, of a community that thins toward vanishing. The framework cannot dismiss this as mere prejudice without abandoning the very account of morality it has defended, on which the felt significance of a threatened good is a real evaluative datum rather than an error to be corrected by fiat.

The reply grants the datum and locates the confusion, and the distinction it draws is the one the whole chapter turns on. The empathic concern for a community's continuity is real and morally weighty; what does not follow is the evaluation of the prospective partner by creed. The two come apart as empathic evaluation and orthotic assessment come apart (Chapter 16). A parent may grieve the anticipated attenuation of a tradition, an empathic response to a real loss, without judging the person their child loves to be, on account of his creed, less worthy, less good, or less fit to be loved.

The wall is therefore not the grief. It is the transfer of the grief onto the person, so that his standing is fixed by his allegiance rather than his conduct. A tradition can pursue its own continuity by every means deep persuasion permits — by teaching, by drawing the partner in through invitation rather than condition, by making the form of life worth transmitting — and violate nothing the analysis identifies. It crosses into the divine wall only when continuity is secured by evaluating the outsider as inferior for being an outsider, which is the point at which an empathic concern has been made to license an orthotic judgement of persons.

So the objection, correctly handled, refines the analysis rather than overturning it. The survival concern is admitted as real and even honoured; what is denied is that it warrants the creed-based evaluation of the person. The empirical bridging effect (Putnam and Campbell) suggests, moreover, that the traditions which distinguish the two — welcoming the partner while sustaining the practice — are not the ones that dwindle. On the same evidence, they are the ones best able to transmit a faith worth choosing under conditions of freedom.

The interfaith test case reveals the divine wall's deepest consequence: it obstructs the dissolution of creed-based boundaries through the most intensive form of contact the empirical record identifies, the shared life of an interfaith family.

3.5 The comparative evidence

The comparative literature. The doctrinal-gradient prediction should also be placed against the comparative literature on religion and democracy, because that literature has already tested, and embarrassed, the cruder claims in the neighbourhood, where the difference between those claims and this chapter's is what the placement makes visible.

The crude claim is that religiosity as such depresses democratic capacity, and the data are unkind to it. Highly religious societies sustain durable democracies, the United States and India being standing cases, and survey evidence from the World Values Survey shows publics across the major traditions, Muslim publics included, endorsing democratic ideals at comparable rates, with the measurable divergences concentrated on gender and sexuality rather than on democracy itself (Inglehart & Norris, 2003).

The specific findings sharpen the picture. Fish (2002) found a strong association between Muslim-majority status and authoritarian government; Stepan and Robertson (2003) showed the gap to be Arab rather than Muslim once the two are separated; Donno and Russett (2004) re-examined both results and found the linkages running chiefly through the status of women. Stepan's (2000) broader finding frames the field: democracy requires "twin tolerations", religious authority tolerating democratic lawmaking and democratic authority tolerating religious expression, and traditions across the spectrum have historically constructed both.

The framework's prediction is the doctrinal gradient of §3.4, so the comparative record is what that gradient expects rather than what it must explain away. The thesis indexes the deficit to creed-weighting — to the evaluation of persons by allegiance, backed by infallibility and authority-dependence — and creed-weighting cross-cuts religiosity.

A highly observant society whose traditions weight conduct over creed is therefore predicted to be democratically capable, where the standing counterexamples to the crude claim are of just this kind. The religiosity the United States sustains is denominationally plural and competitive, which structurally weakens any single creed's claim to condition civil standing, while the Indian case combines intense observance with largely orthopractic traditions. Stepan's twin tolerations name, in institutional vocabulary, what the second and third transformations of §4 derive structurally: a tradition that accepts revisable law and does not condition standing on allegiance has already made the moves the analysis identifies. The comparative literature therefore constrains the thesis without deciding it, because it rules out religiosity as the operative variable, which is what the framework claims; it has not yet measured doctrine at the grain the gradient requires.

The confounds, specified in advance. The interfaith-marriage test of §3.4 is where that grain can be had, and its confounds should be specified in advance, since a gradient not isolated from them decides nothing. Three matter most.

Endogamy can be ethnic rather than doctrinal, a boundary of peoplehood rather than of creed, so the test requires comparisons within an ethnicity across doctrinal weightings, and across ethnicities within one. Endogamy can also be a survival strategy of small or historically threatened communities, the concern §3.4 has already distinguished from the wall, so group size and minority status must be controlled. And religious sorting tracks socioeconomic sorting, so apparent doctrinal effects must survive controls for education, income, and geography.

A fourth confound survives those three and must be met by design rather than by control: boundary maintenance as such. A community may guard its boundary for reasons that owe nothing to creed — continuity, kinship, the reproduction of a way of life — and closure of that kind is indistinguishable from creedal weighting in any survey that measures only the boundary's strength. The discriminating comparison is therefore between traditions of similar closure that differ in creedal weighting: an orthopractic community and a credal one, each strongly

endogamous, matched on size, minority standing, and ethnicity. The wall predicts that private intolerance thins under interfaith acquaintance in the first and persists in the second, whereas closure alone predicts no difference. The same comparison can be run within one tradition, across congregations of one ethnicity that differ in the weight they place on standing before God; it is there, rather than across whole traditions, that the gradient is cleanest to read.

The prediction, stated with its controls: among communities matched for ethnicity, size, and socioeconomic position, resistance to interfaith intimacy — and the intergroup suspicion it indexes — should track the creedal weighting of the tradition, and should persist among the doctrinally committed at high levels of contact. A flat result under those controls falsifies the divine wall as a doctrinal structure; a gradient confirms it. The chapter stakes the thesis there.

The stake can be made actionable, because data adequate to a first approximation already exist. The Pew Research Center's Religious Landscape Studies (Pew Research Center, 2015) carry intermarriage measures by tradition; the General Social Survey carries religiosity and social-distance batteries under demographic controls; the Faith Matters survey behind Putnam and Campbell's analysis records interfaith ties alongside doctrinal commitment. A first-pass gradient can be estimated from these under partial controls, while the full test, with the within-ethnicity comparisons specified above, awaits data collected at that grain.

Where the argument stands: four convergent results have been assembled (§2), the divine wall has been derived and its most compassionate defence answered (§3), and the thesis has been staked on a doctrinal gradient that can fail. What follows derives the transformations a tradition would have to make.

4 Three Structural Transformations

The divine wall is not inevitable. It is a consequence of specific doctrinal features that are, in principle, removable — structural in the sense that the features generate it wherever the projection of §3.2

occurs, and empirical in whether they do, which is what §3.5 stakes. The three transformations that follow reduce, when their derivation is traced, to the acceptance of Chapter 16 by a tradition's own resources; the reform voices cited are contested minorities; what the section adds is the structural reason their resources point where they do. This section derives three transformations and engages with internal reform movements within the traditions that have already moved in each direction.

4.1 From theocratic to humanistic religiosity

In theocratic religiosity, God is more important than humanity, since He is ontologically superior and devotion to Him takes priority over love for persons. This hierarchy is the root of the divine wall: if God outranks persons, then allegiance to God outranks conduct toward persons, and creed outranks character.

The first section of this book established that no mind can occupy the foreclosing ground of being (Chapter 10). A God who is a mind therefore stands within the hierarchy of diversity, either as the sustainer on whom all else depends or as a companion who shares existence (Chapter 24, §4.1). On neither reading does the theocratic ranking follow, for the reason Chapter 16 gives.

Morality is conduct's bearing on those it affects. A God who feels holds standing as an affected party just as persons do (Chapter 22, §4): graded by what is done to him, and by nothing he made or sustains. Devotion to God and care for persons are therefore two instances of one evaluation, so a framework that ranks the first above the second has replaced the empathic evaluation that makes either moral with an orthotic test of allegiance. The transformation follows: humanistic religiosity — devotion to God expressed through love for persons rather than in competition with it — is the structurally coherent form.

Internal reform movements have already moved in this direction, and three worked cases show the structure. Within Islam, An-Na'im (2008) argues that Islamic values are best served by a secular state, since the state's coercive imposition of Islamic law corrupts the very

values it purports to enforce — theocratic imposition being enforcement rather than deep persuasion. He argues further that the incompatibility of state-applied Sharia with citizenship and equality is recognised within the Islamic scholarly tradition itself, which has always distinguished the eternal principles of the Quran from their historically contingent applications.

His structural move, separating the principles from their applications and subjecting the applications to revision, is the distinction between the empathic foundation — the principles of justice and compassion the revelation embodies — and the orthotic superstructure, the legal codes derived from them in particular contexts: the foundation preserved, the superstructure made revisable.

Within Christianity, Volf (2011) argues that the public expression of faith must be governed by the recognition that Christians share civic space with non-Christians, and that this recognition is a consequence of the Christian commitment to love of neighbour rather than a concession to secularism. The analysis provides the explanation, since the love commandment is the empathic foundation and creed-based evaluation an orthotic distortion the foundation, correctly understood, repudiates. His engagement with pluralism accordingly treats non-Christians as persons of equal moral standing rather than as souls to be saved, and liberation theology reaches the same ground from the other side, locating devotion in service to the marginalised and the suffering (Gutiérrez, 1971/1973).

Within Judaism, Borowitz (1991) develops a covenantal theology in which the covenant between God and Israel is a living relationship rather than a static set of commands. It evolves as the Jewish people's understanding of God's will deepens, so human moral insight is a partner rather than a mere recipient. On the present analysis, Borowitz preserves the empathic core while making the orthotic layer responsive to each generation's evaluative engagement.

The Talmudic tradition's own internal mechanisms of interpretation are structural resources for the second transformation: the principle of makhloket l'shem shamayim, dispute for the sake of heaven, and the

recognition that both positions in a halakhic dispute may be "words of the living God" (Eruvin 13b). A tradition that recognises legitimate disagreement as a vehicle for deepening understanding has already internalised the structural conditions for deep persuasion within its own interpretive framework. The analysis therefore does not ask such traditions to import alien values; it asks them to extend the interpretive openness they already practise internally to their engagement with the broader democratic community.

These movements show that the transformation from theocratic to humanistic religiosity is not imposed from outside. The traditions contain the resources. Two qualifications apply. The engagement above is with each figure's structural move — the distinction each draws between a foundation and its revisable application — and not with the whole of any position, so the appropriation is by the framework's lights and each figure would state the move in his own. And these are contested minority positions: An-Na'im and Abou El Fadl within Islamic scholarship, Borowitz from Reform rather than Orthodox Judaism, Volf from progressive rather than dominant Christianity. Their minority status does not diminish their structural relevance, since the point is that the traditions contain these resources; whether the traditions follow these voices is a matter of deep persuasion within each, and the framework's contribution is the structural explanation of why the resources point where they do.

4.2 From absolutist to democratic religiosity

In absolutist religiosity, scripture is treated as infallible, so it becomes an unreformable standard that cannot be challenged, revised, or corrected by human insight. This is the four-stage anatomy of dogma (Chapter 17, §6) in its home case — empathically earned insight codified, extended beyond its warrant, hardened, and severed from the foundation that gave it authority. What the fossilisation costs is the ratchet (Chapter 17, §7.1), since the layer that would bank the community's empathic corrections is converted into the layer that holds them out.

The transformation follows. Democratic religiosity treats its moral traditions as the current best, subject to revision through future persuasion as democratic law is. That is the perfection/infallibility distinction (Chapter 17) applied to the tradition's own standards. The community's moral insight is preserved while its claim to infallibility is relinquished — the ratchet restored to forward operation, the anatomy run in reverse.

Within Islam, Abou El Fadl (2001) distinguishes the authoritative, the scholarly tradition's careful engagement with the sources, from the authoritarian, the imposition of a single interpretation as permanently binding. He argues that the Islamic tradition's own principles — shura (consultation), ijtihad (independent reasoning), and ikhtilaf (legitimate disagreement) — support a democratic rather than an absolutist approach to religious law.

Within Christianity, the historical-critical method of biblical scholarship, practised by theologians within the traditions rather than imposed from outside, treats scripture as a historically situated text whose meaning is enriched by critical engagement. Sachedina (2001) provides a systematic account of Islamic foundations for pluralism. He argues that the Qur'anic concept of fitra, the innate disposition toward God shared by all human beings, grounds a form of merit-based evaluation available within the Islamic tradition itself. On Sachedina's reading, the Qur'an's emphasis on righteous conduct — "compete with one another in good works", 5:48 — provides the resources for a merit-based evaluation that calls for the deepening of Islamic identity rather than its abandonment. The movement is from a creed-centred to a conduct-centred understanding of what God requires. Within Judaism, the makhloket tradition already cited (§4.1) does the same work here, since interpretive disagreement is treated as a feature that keeps the tradition alive rather than a defect.

4.3 From creed-based to merit-based evaluation

If morality is constitutively empathic, then the morally relevant criterion for evaluating persons is their character rather than their creed:

how they treat others, whether they alleviate suffering, whether they promote flourishing.

The transformation follows: if persons are evaluated by character and not creed, the divine wall dissolves. A non-believer who treats others with compassion is morally superior to a believer who does not, regardless of their respective allegiances. Merit-based evaluation — judging persons by their empathic conduct and not their doctrinal conformity — is the structurally coherent form of moral evaluation within any religious community.

The merit-based direction is already present within the traditions. The Quranic verse “the most honoured of you in the sight of Allah is the most righteous of you” (49:13) grounds divine evaluation in taqwa (God-consciousness expressed through righteous conduct), not in tribal or doctrinal identity. The Christian parable of the sheep and the goats (Matthew 25:31–46) evaluates persons by their conduct toward “the least of these” — the hungry, the stranger, the prisoner — not by their doctrinal confession. The Jewish emphasis on tikkun olam (repairing the world) grounds moral obligation in the practical work of making the world more just, not in credal conformity.

The analysis does not claim that these traditions have fully implemented merit-based evaluation, since many have not. It claims that the traditions’ own deepest resources point toward merit-based evaluation, so the creed-based overlay is the structural anomaly rather than the merit-based foundation.

5 What Remains for Religion

The three transformations remove specific doctrinal features: the ontological superiority of God over humanity, the claim of scriptural infallibility, and the primacy of creed over character. They are independently motivated and can be adopted incrementally, since a tradition could adopt one without the others — moving from absolutist to democratic religiosity while retaining a strong creed-based identity, for example, or abandoning scriptural infallibility while preserving a traditional God-concept. The analysis identifies the direction of reform

without requiring that every tradition arrive at the same destination simultaneously. What remains after the transformations is substantial.

Faith in a God of infinite diversity remains coherent and untouched by the structural arguments: a sustainer on whom all else depends, or a companion who accompanies humanity through existence (Chapter 24, §4.1), powerful, knowledgeable, and loving on either reading. The next chapter develops the second of these in full.

Institutional religion retains its value as a source of community, contemplation, ritual, moral education, and empathic formation. The practical resources of religious life — communal worship, practices of compassion, traditions of reflection on suffering and flourishing — work because they engage the evaluative modality, so they can be preserved without the doctrinal features the transformations remove.

C. Taylor (2007) provides the most comprehensive account of how religious life can persist and deepen in a secular age. The "conditions of belief" have changed irreversibly, since faith is now one option among many, and a faith chosen under real pluralism may be deeper than one held by social default, because it is constituted by primary value, direct engagement with the transcendent, rather than by derived value, the authority of a monopolistic tradition. The three transformations are compatible with Taylor's account, because they remove the derived-value elements while preserving, and potentially deepening, the primary-value engagement he identifies as the authentic core of religious experience.

Contemplative practice — prayer, meditation, reflective engagement with sacred texts — retains its value as a means of deepening the modality's capacity to detect evaluative differences, since a practice that heightens empathic sensitivity is a real resource for the moral life whether or not the practitioner believes the contemplative object is a creator God. Moral education within religious communities retains its value insofar as it cultivates empathic engagement rather than orthotic compliance. What is lost is the claim that moral authority derives from divine command; what is gained is the freedom to revise

moral teachings in light of deeper empathic insight, a freedom infallibility claims foreclose.

What is removed, then, is the claim to absolute authority rather than faith. What is preserved is the evaluative engagement with the transcendent that gives religious life its depth, meaning, and beauty. The question is not whether religious life survives the structural critique, since it does, substantially, but whether the structurally incoherent elements are essential to what believers love about their traditions, or whether what they love can be preserved, and perhaps deepened, once those elements are set aside.

6 Objections and Replies

6.1 The blasphemy objection

Objection. The chapter asks religion to give up its most fundamental commitments — God's supremacy, scriptural authority, and the salvific importance of faith — so the "transformations" amount to asking believers to abandon their faith.

Reply. The chapter asks religion to give up three specific claims: that God is ontologically superior to humanity in the foundational sense the theocratic framework requires, that scripture is infallible, and that creed determines moral worth. These are distinct from faith in God, love for God, or the value of religious community, since a person can maintain a loving and profound relationship with God while recognising that God did not produce being from its absence, that scripture is humanly mediated, and that persons should be evaluated by character.

The internal reform movements of §4 show that believers have already made this distinction. Liberation theology's preferential option for the poor is grounded in empathic engagement with the suffering of the marginalised rather than in creed-based evaluation of their doctrinal correctness (Gutiérrez, 1971/1973), and progressive theology more broadly, from Tillich's (1952) "God above God" to the process theology of Chapter 24, has consistently moved in the direction the analysis identifies.

6.2 The secularism objection

Objection. The chapter systematically dismantles religious claims while leaving secular ideologies unchallenged.

Reply. The analysis challenges any ideology, religious or secular, that disables the conditions for democratic peace. Wolterstorff (2012) argues that the Rawlsian demand for "public reason" is itself a form of secularist exclusion, while Stout (2004) argues that democratic deliberation requires citizens to translate their religious convictions into reasons their interlocutors can engage with rather than to bracket them. The framework is closer to Stout, since deep persuasion requires that all evaluative hierarchies, including religiously grounded ones, enter the deliberative field — because persuasion engages the parties' actual commitments rather than sanitised substitutes. Habermas's (2006) institutional translation proviso is the intermediate position: religious reasons admitted in the informal public sphere and translated into generally accessible terms at the legislative threshold. The framework's line falls where Habermas draws that threshold, since the field must admit the conviction as it is held, and the law codifies only what the field has come to share.

What it adds is the structural diagnosis. The problem is not that religious reasons enter deliberation, since they should, but that the claim to infallible authority disables the deliberation's capacity to produce transformation: the citizen who brings her empathic convictions is engaging in deep persuasion, while the one who asserts infallible authority forecloses persuasion altogether. The previous chapter argued that moral relativism is as structurally destructive as moral absolutism, and the divine wall is targeted because religious absolutism is, empirically, the most socially consequential form. The analysis is democratic.

6.3 The reform-from-within objection

Objection. Religious traditions have always reformed from within. The chapter's transformations are unnecessary because the traditions already have the resources.

Reply. The chapter agrees (§4). The three transformations are not imposed from outside; they are derived from structural results that the traditions can engage with on their own terms. The chapter's contribution is to provide a structural foundation for these movements — to show that the direction of reform is a structural consequence of what the preceding chapters have established.

6.4 The scope objection

Objection. The chapter claims to target "traditions in which creed-based evaluation is primary" but many of its arguments apply more broadly.

Reply. The qualification is a matter of degree, since no sharp boundary is implied. The chapter's argument is a structural gradient: the more a tradition weighs creed over character, the more intensely the divine wall operates, and the more the tradition's democratic capacity is structurally compromised. A tradition that already evaluates persons primarily by character has less to revise, while it still benefits from the explanation of why the conduct-centred direction is the correct one.

6.5 The eschatological objection

Objection. The doctrine of divine judgement concerns the afterlife; human social arrangements are another matter. The projection from celestial judgement to earthly division is the chapter's inference, not the tradition's.

Reply. The projection is the tradition's rather than merely the chapter's. When a community believes that God will ultimately separate the blessed from the damned by creed, that belief constitutes a present-tense evaluative framework, since it tells the believer who is on the right side and who is not, right now.

The interfaith marriage case (§§3.4–3.5) is where the thesis stakes itself empirically. The doctrinal gradient — endogamy norms and intergroup suspicion tracking the creedal weighting of the tradition — is consistent with the available comparisons (Putnam and Campbell) though not yet isolated by them, so the projection is falsifiable there,

under the controls §3.5 specifies. That the prediction runs along the tradition's own doctrinal axis is the reply's point, because the gradient tests whether the projection is the tradition's, where the chapter's inference could lose.

6.6 The pastoral concern

Objection. The chapter's analysis is intellectually coherent but pastorally insensitive. For many believers, the doctrines under critique provide comfort, meaning, and structure. Removing them creates a pastoral vacuum the analysis does not fill.

Reply. This objection deserves the most careful response in the chapter, because it identifies a real cost that intellectual analysis alone cannot address. This chapter asks more of its readers than any other in the book. The beliefs under critique are not merely propositions to be assessed; they are the structures within which many people have built their lives, their relationships, and their sense of who they are. Releasing them is not a philosophical exercise but, for many, a form of bereavement. The evaluative modality, the analysis's own concept, explains why: the believer is losing secondary objects, the derived value of traditional associations, that have populated her evaluative hierarchy for a lifetime. The grief is real.

What the chapter also means to show is that the loss is not total. The remedy is what §5 describes: the preservation and deepening of the evaluative engagement that gives religious life its depth. The tradition that emerges may be deeper, not shallower, than the one it replaces — because its evaluative engagement is grounded in primary value and not in derived value that rests on structurally incoherent foundations.

7 Conclusion

The chapter yields three results, all constructed.

The divine wall. The divine wall is the consequence of creed-based evaluation applied to persons — orthotic assessment in its most consequential social form — wherever post-mortem judgement by allegiance is projected onto present standing; that the projection occurs

is staked on the doctrinal gradient of §3.5. It operates at the most private level of intolerance and is structurally resistant to legal enforcement.

Three transformations. Three structural transformations follow as consequences of the established results: from theocratic to humanistic religiosity, from absolutist to democratic religiosity, and from creed-based to merit-based evaluation. Internal reform movements within the major traditions — Volf, An-Na'im, Borowitz, Abou El Fadl, and the Talmudic and Quranic resources cited — have already moved in these directions, for which the analysis provides the structural explanation of why these resources point where they do.

What remains for religion. What remains is substantial, since faith in a God of infinite diversity, institutional community, contemplative practice, moral education, and empathic formation survive intact. Only the claim to absolute authority and the evaluative primacy of creed over character are removed. Removing the doctrinal features the preceding chapters have shown structurally incoherent therefore opens space for a form of theism that preserves what believers value most — community, contemplation, devotion, the experience of the sacred — while dissolving the evaluative barriers creed-based judgement erects between communities.

The next chapter completes the movement by asking the converse question. This chapter argued from the conditions of democratic peace to the forms of religion that obstruct them; Chapter 24 asks which conception of God those conditions leave standing, drawing its answer from the metaphysical results of the book's first half. The two directions meet in a single figure — a God who is the structural negation of the theocratic God examined here, a companion who persuades where the absolutist commands — and in a constructive account of what God and faith could be once the structurally incoherent claims have been removed: what theism looks like after zero complexity.

Grades and Dependencies

1. Conditionality and scope. All three results are *constructed*, and the chapter is conditional on the empathic constitution of morality (Ch. 16): a divine-command theist resists by resisting that premise. The argument is a gradient, strongest where creed-weighting is primary (§§3.1, 6.4). The circularity diagnosis addresses the assessor outside the framework; for the believer within it maps, without persuading (§3.2). The Premise Register carries the chain.

2. Evidence and limits. Putnam and Campbell's findings are correlations; the bridging effect is read as the wall dissolving through contact, kept falsifiable by the gradient test: flat under controls refutes (§§3.4–3.5). Comparative work rules out religiosity as the variable (Fish; Stepan & Robertson; Donno & Russett) without measuring doctrine at that grain. §4's reform voices are contested minorities; their weight is structural, not demographic.

References

Abou El Fadl, K. (2001). *Speaking in God's name: Islamic law, authority, and women*. Oneworld.

Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley.

An-Na'im, A. A. (2008). *Islam and the secular state*. Harvard University Press.

Borowitz, E. (1991). *Renewing the covenant: A theology for the postmodern Jew*. Jewish Publication Society.

Donno, D., & Russett, B. (2004). Islam, authoritarianism, and female empowerment: What are the linkages? *World Politics*, 56(4), 582–607.

Fish, M. S. (2002). Islam and authoritarianism. *World Politics*, 55(1), 4–37.

Gutiérrez, G. (1971/1973). *A theology of liberation* (C. Inda & J. Eagleson, Trans.). Orbis.

Habermas, J. (2006). Religion in the public sphere. *European Journal of Philosophy*, 14(1), 1–25.

Hauerwas, S. (1981). *A community of character: Toward a constructive Christian social ethic*. University of Notre Dame Press.

Hick, J. (1989). *An interpretation of religion: Human responses to the transcendent*. Yale University Press.

Inglehart, R., & Norris, P. (2003). The true clash of civilizations. *Foreign Policy*, 135, 62–70.

Paluck, E. L., Green, S. A., & Green, D. P. (2019). The contact hypothesis re-evaluated. *Behavioural Public Policy*, 3(2), 129–158.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90(5), 751–783.

Pew Research Center. (2015). *America's changing religious landscape*. Pew Research Center.

Plantinga, A. (1995). Pluralism: A defense of religious exclusivism. In T. D. Senor (Ed.), *The rationality of belief and the plurality of faith* (pp. 191–215). Cornell University Press.

Plantinga, A. (2000). *Warranted Christian belief*. Oxford University Press.

Putnam, R. D., & Campbell, D. E. (2010). *American grace: How religion divides and unites us*. Simon & Schuster.

Sachedina, A. (2001). *The Islamic roots of democratic pluralism*. Oxford University Press.

Stepan, A. (2000). Religion, democracy, and the "twin tolerations". *Journal of Democracy*, 11(4), 37–57.

Stepan, A., & Robertson, G. B. (2003). An "Arab" more than a "Muslim" electoral gap. *Journal of Democracy*, 14(3), 30–44.

Stout, J. (2004). *Democracy and tradition*. Princeton University Press.

Stump, E. (2010). *Wandering in darkness: Narrative and the problem of suffering*. Oxford University Press.

Taylor, C. (2007). *A secular age*. Harvard University Press.

Tillich, P. (1952). *The courage to be*. Yale University Press.

Volf, M. (2011). *A public faith: How followers of Christ should serve the common good*. Brazos Press.

Wolterstorff, N. (2012). *Understanding liberal democracy: Essays in political philosophy* (T. Cuneo, Ed.). Oxford University Press.

CHAPTER 24

Peace and God

Presupposes: the disjunctive ecumenical result and the unilateral boundary (Vol. I, Ch. 4); the temporal terminus and the order of time (Vol. I, Ch. 5); the unintelligence of the foreclosing terminus (Vol. I, Ch. 6); the two infinities (Vol. I, Ch. 8); the dissolution of creation ex nihilo (Vol. I, Ch. 9); the ground/mind conflation (Vol. I, Ch. 10); the failure of the kalām's necessary cause (Vol. I, Ch. 11); the categorical transition principle (Ch. 13); the response-constitutive account of value (Ch. 14); the primary/added value taxonomy (Ch. 15); the descriptive mechanism statement (Ch. 16); the divine wall of discrimination (Ch. 23).

Establishes: the two conceptions of God that survive the critique on the diversity side, sustainer and companion; the companion God, a temporal mind and the form of theism compatible with structural peace; and, in the appendix, faith as evaluative orientation and the process-theology convergence.[2]

In brief. When the traditions say "God is infinite," which infinity do they mean? Everything turns on the answer. Infinite equality is the terminus — unintelligent, unconscious, undifferentiated being at zero complexity; infinite diversity, a maximal mind that did not originate existence from nothing. No third condition exists; the classical incoherences come from demanding both at once. Two conceptions survive on the diversity side, a sustainer on whom all else depends and a companion who shares existence without grounding it, and the chapter develops the second: God as companion, not creator — a fellow inhabitant who lures and does not command, whose only power can be freely declined. The proposal sits on the ladder's weakest tier, mindedness and benevolence both costed posits. One trade is owned outright: a God who could guarantee rescue, or one not answerable for its absence — not both.

---

1 Introduction

Chapter 23 showed which forms of religion obstruct structural peace: those that judge persons by creed rather than character, treat their standards as infallible, and rank devotion to God above regard for persons. This chapter asks the converse question. What God, and what form of belief, is compatible with peace, or even conducive to it?

The question must be approached in the right order, or the answer is worthless.

One cannot begin from the peace one wants and posit whatever God would secure it, because that would be wishful theology, and the analysis permits no such move. The order is the reverse. The preceding chapters have already removed six specific theological claims on independent structural grounds — that God created existence from nothing, that the prime mover is a mind, that divine simplicity is compatible with intelligence, that divine retribution is a just response to free choices, that creed-based judgement is a legitimate criterion of moral evaluation, and that scripture is infallible — so the question is what survives their removal. The argument of this chapter is that what survives, on grounds fixed by Chapters 8–12 which owe nothing to the politics, turns out to include the very God that peace allows.

One concession belongs in the same breath, because the chapter's candour depends on it. The metaphysics alone fixes which conceptions of God are coherent, while the choice of which coherent survivor to develop, in a book that has just derived the conditions for peace, is openly guided by those conditions. What the order of argument secures is that the coherence space was drawn before the selection was made, and drawn by arguments the politics never touched.

The answer is: more than the critique might suggest. The argument refutes specific conjunctions — divine simplicity with divine intelligence, creation with zero complexity, moral evaluation with creed-based conformity. The chapter identifies the deep source of these incoherences in the conflation of the two infinities (Chapter 8), the demand that God be both infinitely diverse and infinitely equal.

Once the conflation is resolved, two structurally coherent theisms remain on the diversity side: a sustaining God on whom all else depends, and a God of infinite diversity who is companion rather than creator (§4.1), together with a faith of empathic orientation rather than orthotic obedience. That the companion is also the survivor compatible with peace is a structural consequence rather than a further stipulation, because a companion who lures and does not command relates by persuasion rather than imposition, which is the theological form of the deep persuasion on which peace depends (Chapters 21–22).

The relation between the two inquiries is this chapter's organising thread, and its direction should be stated in full. The metaphysics draws the space of coherent options without consulting the politics. The politics guides which option the chapter develops. The claim is one of selection: the politics chose, from a two-member space the metaphysics had drawn, the member whose only mode of address is persuasion; a persuasion-only God is compatible with a persuasion-based peace by construction. What the order of argument secures is that the selection was made from a space drawn without it, not that the two inquiries met.

The chapter summarises what the preceding analysis removes (§2), locates the deep source of the incoherence in the conflation of two infinities (§3), identifies what it does not remove (§4), develops the concept of God as companion (§5), and meets the objections (§6); faith, worship, and the process lineage, consequences of the proposal, are set out in an appendix (§§A.1–A.8).

2 What the Analysis Removes

The theological critique is narrow in its targets. It does not argue against God in general; it argues against specific claims about God that are structurally incoherent. Each has been argued in full in earlier chapters; this section states them compactly. The critique is also narrower than its vocabulary sometimes suggests: a theist who never held simplicity, timelessness, retributive desert, or creed-based judgement — the open theist, the personalist — finds most of this section aimed past her; what remains for her is the sustainer's cost (§4.1) and the inefficacy and hiddenness residues the companion carries (§§6.5–6.6).

Creation ex nihilo (Chapter 9). Creation pictured as a transition from non-being is incoherent, and the mature doctrine — creation as non ex aliquo, total ontological dependence — is met by the agent dilemma, whose second horn Chapter 9 inherits from Chapters 6 and 10. Either an originating agent is within diversity, in which case its act is diversification rather than origination, or it is at the undifferentiated limit, in which case it has no agency to exercise. What survives is

ontological dependence, which is coherent and which the analysis does not contest; what does not survive is origination from non-being. The sovereignty doctrine loses nothing by that removal (Chapter 9, §6.2), because what it costs is met later, by the companion God alone (§6.5).

Four further removals follow the same pattern. The prime mover as mind (Chapter 10): the cosmological tradition conflates being the ground of existence with being a mind that chooses, roles the analysis shows are structurally incompatible. Divine simplicity conjoined with intelligence (Chapter 10): the classical doctrine that God has no internal distinction is incompatible with knowledge, will, or intention, since all of these require distinction.

Divine retribution as just punishment (Chapter 20): bounded authorship makes absolute moral responsibility unavailable to any conscious agent. Creed-based moral judgement (Chapter 23): evaluating persons primarily by doctrinal allegiance rather than conduct is the great moral conflation in its most consequential social form. Scriptural infallibility (Chapter 17): perfection does not entail infallibility, so the conflation fossilises revisable moral technologies into unreformable dogma.

These claims are the analysis's targets. They are specific and independently argued. The question is whether anything recognisably theistic survives their removal.

3 God and the Two Infinities

Chapter 8 established that the concept of infinity is ambiguous between two structurally opposite conditions. Infinite diversity is the endless extension of internal structure — more distinctions, more complexity, more differentiation, without limit. Infinite equality is the complete dissolution of all structure into undifferentiation — no distinctions, no complexity, zero internal variation. These two conditions are not different degrees of the same thing; they are opposite ends of the structural spectrum.

When religious traditions say “God is infinite,” which infinity do they mean? The entire coherence of the theological concept depends on the answer.

Suppose God is infinitely diverse — endlessly rich in attributes, knowledge, and internal complexity. Then God is maximally differentiated, and, if that diversity is instantiated as detection, maximally conscious, since consciousness is constituted by detected distinction rather than by distinction possessed (Chapter 13: the universe is diverse and is not a subject). Such a God is also maximally evaluative (Chapter 14, because the evaluative modality requires evaluative differences to detect) and maximally capable of purpose (Chapter 19, because purpose requires the evaluative modality’s engagement with a field of differences).

The conditional’s antecedent is supplied by the proposal of §5 rather than by the concept of diversity. This God is coherent with the present analysis’s results, though also maximally complex: the structural opposite of the classical doctrine of divine simplicity. An infinitely diverse God has infinite internal distinction, which is what divine simplicity denies.

Suppose instead God is infinitely equal — a state of zero complexity, complete undifferentiation. Then God is the grounding terminus itself. But the terminus is unintelligent (Chapter 6), unconscious (Chapter 13), without evaluative capacity (Chapter 14), and incapable of purpose (Chapter 19). This God therefore lacks personhood, mentality, and the capacity to love, judge, or create. This God is undifferentiated being at zero complexity.

The conflation of the two infinities is the deep source of the six incoherences identified in §2. Creation ex nihilo requires a God at the terminus (infinite equality) who possesses intelligence and will (infinite diversity). The prime mover tradition places God at the ground of existence (infinite equality) while attributing mental properties (infinite diversity). Divine simplicity identifies God with undifferentiation (infinite equality) while claiming he knows and chooses (infinite

diversity). In each case the tradition demands both infinities at once, and the analysis shows that the two are structurally incompatible.

The resolution is to choose. Either a God of infinite diversity who did not originate existence from non-being, or a terminus of infinite equality that is not God; no third condition exists for a state to be in. The first horn itself divides, and §4.1 draws the division: a God of infinite diversity may be the ground on which all else depends, or a fellow inhabitant of existence. The constructive proposal that follows chooses the first horn, and within it the second figure.

4 What the Analysis Does Not Remove

4.1 Two survivors: the sustainer and the companion

The argument does not refute the existence of a being of immense, even infinite, diversity: a being of extraordinary power, knowledge, and love. What it rules out is that such a being originated existence from non-being (Chapter 9). Two figures survive on the diversity side, and the ledger of the preceding chapters leaves both standing.

	The sustainer	**The companion**
Relation to existence	The ground on which everything else depends: creation read as dependence without origination (Chapter 9, §2.3)	A fellow inhabitant of existence, who finds himself within it as we do
What the earlier chapters allow	A complex ground escapes the zero-complexity constraint (Chapter 6, §6.4); Chapter 10 targets only the simple ground; Chapter 11 shows only that such a being cannot be shown necessary	A being within the hierarchy of diversity, bounded by the terminus as every complex being is (§4.2)

	The sustainer	**The companion**
Sovereignty	Kept: nothing other than God stands on its own	Given up: he sustains nothing and originates nothing
What he answers for	The causal order he holds in being; the problem of evil returns at full strength, and the educative argument of Chapter 20 (§6) presses on him	Nothing he lacks the power to prevent; his one mode of address is persuasion, which leaves the other free (§6.5)
Lineage	Open theism (Hasker, 2004; Ward, 2007); the classical God once the virtual distinctions of his knowing are counted as contrasts (Chapter 10, §5.1)	Process theology's consequent nature (§A.6)

The two differ in what each must answer for, and the difference grounds the chapter's choice. The coherence space was drawn by Chapters 8–12 without consulting the politics, while the selection within it is guided by Chapters 21–22, whose conditions for structural peace are conditions a persuader satisfies and a sustainer need not. The chapter accordingly develops the companion, leaving the sustainer named as the live alternative, with its cost itemised where it falls (§4.2, §6.5). The companion belongs among the alternative concepts of God the recent literature has catalogued (Buckareff & Nagasawa, 2016), as a personal non-creator theism distinct from the pantheist and panentheist members of that family, since the companion is neither identical with the world nor its ground.

4.2 The ontological status of the companion God

The companion God raises ontological questions that classical theism does not face, since the sustainer inherits the tradition's answers

to most of them along with the tradition's grounding role. The chapter must therefore address them directly rather than leaving them to the reader's inference.

What is this God's relationship to the grounding terminus? The terminus is the structural boundary at which all distinction ceases: the state of zero complexity. God, as a being of infinite diversity, stands at the top of the containment order (Chapter 4, Appendix), as far from the terminus as any state can.

The terminus performs no grounding, because it is a limit-ground only (Chapter 4, §3), so nothing is grounded in it, God included. What it does is bound the hierarchy of diversity on one side, and the companion God is within that hierarchy, bounded as every complex being is. This is a departure from the entire classical tradition, which places God at the foundation. On the analysis's account the foundation is unintelligent, so a God who is a mind lives above it: the sustainer as the first structured member on which the rest depends, the companion as one member among the rest.

Necessity and contingency. Is this God necessary or contingent? Leftow (2012) argues that God's necessity is metaphysical, since God's nature is such that non-existence is not a real possibility for a being of that kind. Pruss (2006) connects divine necessity to the principle of sufficient reason, because only a necessary being can serve as the ultimate explanation for contingent ones. Ward (2007) develops the "open theology" tradition, in which God is really affected by the world's development: a responsive being whose experience is enriched by engagement with creation.

Ward's open God is the sustainer of §4.1 — a creator-sustainer, temporal, responsive, and affected by what happens — so it converges with the companion God on everything except the grounding role. Where Ward derives his concept from the Christian tradition's own resources, the suffering God of the cross and the responsive God of the prophets, the analysis derives the shared features from the resolution of the two-infinities conflation. The convergence of independent

derivations strengthens both, while the divergence marks the line §4.1 draws.

The analysis engages both. If God is within the diversity hierarchy rather than at its foundation, then neither Leftow's metaphysical necessity nor Pruss's explanatory necessity applies in its original form. Leftow's argument presupposes that God is the kind of being whose nature excludes non-existence; but if God is a being within diversity, God's nature is constituted by the structural conditions that produce complex beings, and those conditions may or may not necessitate a being of infinite diversity. Pruss's argument presupposes that the chain of explanation terminates in a necessary being; the analysis agrees that the chain terminates, though the terminus is unintelligent equality rather than God.

The companion God may inherit a form of conditional necessity: if the structural conditions for infinite diversity are satisfied, a being of that complexity may be structurally inevitable. The clause is a large modal claim and should be sized. What could make maximal complexity inevitable is a filling principle: that unbounded diversity, given enough of whatever the hierarchy runs on, saturates every structural niche including the maximal one. The framework asserts no such principle. It records the conditional as an open possibility whose antecedent no argument here discharges.

Even granted whole, the result is weaker than the absolute necessity both Leftow and Pruss defend. This is the conclusion Chapter 11 reached against the kalām's necessary cause, and for the same structural reason: a necessary being of infinite diversity would carry its distinction into every possible world, leaving equality — undifferentiated being — impossible, and that impossibility cannot be established. The necessity classical theism needs is therefore unavailable to a being within the diversity hierarchy.

Emergence. And where did this God come from? Because the companion God is within the hierarchy of diversity rather than at its foundation, the question of emergence is unavoidable. On the non-termination branch of the temporal disjunction (Chapter 5), the question

does not arise: God, like all diversity, has always existed, an infinitely complex being whose history extends indefinitely into the past.

On the termination branch, the transition from equality to diversity is actual, established retrospectively by the existence of diversity, while the question of why it occurred is structurally foreclosed, since equality has no determination and therefore no features from which an explanation could be constructed (Chapter 5, §2.5). The argument does not explain which particular forms of diversity emerge. The existence of a being of infinite diversity is, on the termination branch, either a brute fact or a consequence of some structural principle the preceding analysis has not identified. This is a concession stated in the open: the companion God is not produced by the chapter's argument, but shown by it to be structurally coherent.

This concession must be distinguished from the explanatory cost Chapter 4 charged against a brute terminus, or the framework would be exempting itself from its own standard. The two liabilities are not the same. Chapter 4 penalised a terminus whose internal distinctions are brute: a ground that halts the regress while leaving unexplained the very structure that does the grounding.

The companion God is no such ground. He is an inhabitant of the diversity hierarchy rather than its floor: like every complex being he is bounded by the same equality terminus, and he closes no explanatory regress that the unintelligent terminus has not already closed. What remains open is only which forms of diversity the transition yields, a question the framework holds to be foreclosed from the equality side for all diversity alike (Chapter 5, §2.5) rather than a licence claimed specially for God.

The companion God is therefore brute in no sense Chapter 4 forbids. The ledger Chapter 4 kept against Schaffer should be run here too, or the framework exempts itself: Schaffer's cosmos was charged for the unexplained structure of a terminus that grounds, while a maximal mind that registers the cosmos carries more structure than the cosmos does. The difference the proposal rests on is that the companion grounds no other being, so his structure is not the structure of a terminus

and closes no regress; what remains charged is the bare existence of a maximal mind among the forms diversity took, the largest unexplained fact any inhabitant of the hierarchy could be; the proposal carries it at that size: his existence is contingent as any complex being's is, rather than brute as a self-standing diverse ground would be, which is the sustainer's standing and the sustainer's cost.

Classical theism holds that God is a necessary being, one whose non-existence is impossible. The companion God is not obviously necessary in this sense. If God is within the hierarchy of diversity and not at its foundation, God's existence may be contingent, dependent on the structural conditions that produce beings of extraordinary complexity. The argument does not settle this question; it identifies the structural constraint.

A God whose existence is contingent is still a being worthy of devotion, since the contingency of a relationship does not diminish its value. But the contingency has a further consequence that turns it from a weakness into a relational strength: the relationship between humanity and God is bidirectional. A necessary God who could not fail to exist is not enriched by human engagement, whereas a contingent God who shares existence with humanity is. The companion God's evaluative field is really affected by human suffering and flourishing, as a structural feature of shared existence rather than as a design flaw. The companion God is a being one is fortunate to encounter.

How does this God relate to the causal-temporal hierarchy? The temporal analysis (Chapter 5) established that the terminus is the temporal boundary: the point before which no sequence extends. A God of infinite diversity exists within time, not outside it. This God has a history: a temporal sequence of experiences, engagements, and responses. This God is not eternal in the classical sense (timeless, outside the temporal sequence) but in the diversity sense (having always existed within time, as an endlessly complex being whose history extends indefinitely into the past). The distinction matters: a timeless God cannot respond to events in time (response requires temporal sequence); a temporally extended God can respond, engage, and be affected by what happens in the world.

Where the argument stands: six theological claims have been removed on independent grounds (§2), the deep source of the incoherence has been located in the conflation of the two infinities (§3), and two coherent theisms have been found to survive on the diversity side (§4). What follows develops the one the chapter selects.

5 God as Companion, Not Creator

Here the analysis shifts from critique to construction. The preceding sections established the constraints any coherent theism must satisfy; the account that follows develops a conception of God that meets them. It is offered as a proposal, the weakest tier on the book's confidence ladder, and is to be weighed as such.

5.1 The structural concept

If God did not create existence, what is God's relationship to it? A specific answer emerges: companion. J. A. Keller (2007) argues that divine action in the world need not presuppose creation, since God can act within a world whose existence is independent of him, through persuasion, inspiration, and empathic engagement with conscious beings.

Keller's account of non-creative divine action bears directly on the companion God. If God did not create the universe, the question of how God acts within it becomes pressing, and Keller provides the philosophical resources for an answer that does not depend on the creator model. On this analysis, the companion God's mode of action is deep persuasion: offering evaluative perspectives that conscious beings are free to engage with or decline, without the coercive authority that creation would confer.

Such a God shares existence with humanity. He finds himself within the universe, having not brought it into being, alongside the beings whose suffering and flourishing he witnesses and to which he responds. His relationship to humanity is therefore that of fellow inhabitant to fellow inhabitant rather than that of maker to product.

The political sequence gives the concept its sharpest foil. Chapter 22 diagnosed absolutism as the placement of moral law beyond human reach, and Chapter 23 traced the theological form of that placement: a God who separates souls by creed and thereby licenses the evaluation of persons by allegiance. The companion God is the structural negation of that figure, point for point. Where the absolutist God commands, the companion persuades. Where the theocratic God judges by creed, the companion responds to what persons do and undergo. Where the divine wall separates communities by allegiance, the companion's registration extends to every being that feels.

The bookkeeping of §1 applies here too: the concept was fixed by §3's option-space before any comparison with the theocratic figure was made. A God confined to persuasion is constitutionally incapable of tyranny, since enforcement is the one mode of address unavailable to him, and this incapacity — named in §6.5 as the proposal's gravest cost — is also what fits him to be the theological counterpart of the deep persuasion on which structural peace runs.

This shift has profound consequences for the character of the divine, and the conceptual tools developed in earlier chapters permit a more rigorous derivation than the metaphor of companionship alone provides. A creator God stands above creation: he designed it, he owns it, he judges it against the blueprint. A companion God stands within creation, sharing the same condition as the beings within it.

The categorical transition principle (Chapter 13) supplies the conditional the proposal collects: a being whose infinite diversity is instantiated as detection would possess consciousness of maximal complexity, every contrast available in the hierarchy of mental states instantiated across time, producing an evaluative field of unbounded richness. That the diversity is so instantiated — that it composes a mind rather than an infinitely structured non-subject — is the proposal's first posit, costed here as its benevolence is costed in §5.2. Infinite diversity fixes what a maximal mind would register, not that the diversity composes a mind. The Chapter 13 results are conditionals the proposal collects rather than derivations it performs.

What such a being registers. Such a being would register differences that finite minds cannot. Not because he transcends the structure of consciousness — transcendence of structure is infinite equality, and the companion God is infinite diversity — but because his consciousness instantiates the structure to a degree no finite being approaches.

The evaluative modality (Chapter 14) specifies how this consciousness engages. The companion God's emotional responses are evaluative registrations of the differences he encounters: empathic responses to suffering, evaluative engagement with flourishing, aesthetic sensitivity to structural complexity. A being of infinite diversity does not merely know that suffering occurs; he evaluatively registers it with the full weight of an unbounded evaluative field. This is something both more limited and more profound than classical omniscience, which is propositional knowledge of all truths: it is maximal evaluative engagement with every detectable difference.

The companion God may therefore lack propositional knowledge of future contingents, since such knowledge requires temporal structure that has not yet obtained, while possessing evaluative sensitivity that exceeds any finite capacity. He is affected by human suffering as a being who empathically registers the suffering of those alongside whom he exists, rather than as a designer troubled by a flaw in his product. He responds to human flourishing as a being whose evaluative modality is engaged by it, rather than as a master pleased by obedient servants.

The interface condition. One condition from Chapter 13 applies to this being, and the proposal cannot collect that chapter's conditionals selectively. Chapter 13 (§A.4) locates experience at an interface: an activation standing against the modality's absence, the blind field for vision, silence for hearing. A mind with every contrast present at once would have no absence for any activation to stand against, so on the book's own account it would be a structured non-subject — infinite diversity instantiated, and nothing registered.

The companion God meets the condition because he is in time (§4.2). His registrations are diachronic, so each stands against what he

had not yet registered, so the absence the interface needs is supplied by the not-yet. Infinite diversity of content across time is compatible with finite change at any moment: an unbounded field, fed one difference at a time. Unbounded in reach means reach across time: what the temporal reading rules out is a field that holds everything at one moment, and it permits a field that, fed without limit, leaves no instance permanently outside it — and "maximal registration," wherever the chapter uses the phrase, means the second. The reading of "infinite" the proposal draws on is limitlessness by endless extension (Chapter 8, §4), and that reading alone.

This is what "responsive" means for such a being. It is also why his is the one form of unending life Chapter 20 (§3) found immune to saturation, since inexhaustible novelty is a temporal notion and a mind that registered everything at once would have nothing left to register. The mindedness posit is therefore the posit of a temporal mind, and "every contrast available in the hierarchy of mental states" describes the field's reach across time, each contrast arriving against the absence that preceded it.

The companion God lures and does not command. He offers an evaluative perspective that others are free to engage with or decline. His authority, if he has any, is the authority of a perspective that earns assent through its quality — the authority of deep persuasion, not of enforcement or shallow persuasion. This is a God who operates within the democratic framework established in Chapter 22: he participates in the evaluative exchange as a being of extraordinary capacity, but his participation leaves room for the participation of others.

5.2 The direction of the divine current

One attribute of the companion God has so far been carried by the description rather than the derivation. The difference must be made explicit, because everything pastoral in this chapter and in Chapter 20 hangs on it. Infinite diversity fixes the capacity of the divine evaluative modality: a being instantiating every category of detectable distinction registers, across time and with no instance permanently beyond its reach (§5.1), every instance of suffering and flourishing with maximal

accuracy and completeness. It does not, by itself, fix the direction of the current that rides that registration. Chapter 13 (§2.5) made empathy and cruelty one capacity run in opposite directions; Chapter 14 (§6.4) conceded the coherent outlier — a being that registers another's suffering with full clarity and is drawn to it. A maximally registering being of inverted current is therefore not excluded by the concept of infinite diversity; a derivation of benevolence from structure alone is not available. The chapter says so rather than borrowing the conclusion.

What can be derived is narrower, and worth deriving. The familiar engines of malevolence are, on the framework's own analysis, artefacts of limitation. Cruelty as humans practise it is overwhelmingly the added-value override of a primary empathic response. The in-group's derived value outweighs the out-group member's registered suffering; threat and rivalry convert another's loss into one's own gain; scarcity sets one welfare against another (Chapters 15, 16, 21). Each engine presupposes a bounded position: something to defend, something to lose, someone to compete with. A being of infinite diversity occupies no such position. It has no out-group, since every perspective is within its registration; no rival, since it competes for nothing; no scarcity, since its engagement with one being subtracts nothing from its engagement with another. In such a being the distortion-based routes to cruelty are structurally idle. What remains possible is the pure outlier of Chapter 14: inverted current without misregistration. Against that possibility the framework offers no proof, only the observation that the inversion would be a bare further fact, motivated by nothing the being's situation supplies.

Benevolence is accordingly posited within the proposal. The companion God is proposed as a being of infinite diversity whose current runs with the registration, toward the flourishing of what it registers; the universalism Chapter 20 (§6) proposes is this current at its longest reach — every soul turned toward reconciliation by persuasion across time, an aim such a God keeps and an outcome he cannot guarantee (§6.5). The posit is tiered with the proposal it belongs to, not smuggled in as a derivation. A reader who withholds the posit retains a weaker but still pointed result: the God that survives the analysis is a

being whose moral character structure does not fix. That is itself a finding against the classical tradition, whose central ambition was to derive perfect goodness from perfect being. The framework locates goodness where Chapter 14 locates all value, in the direction of a current; and directions, even divine ones, are not entailed by capacities.

One further cost belongs beside the posit, because the derivation just given has a shape the framework elsewhere polices. Reading the divine motivational field through the human engines it excludes is an inference of the projection kind (Chapter 8, §3). The engines audited are ours; their absence in an unbounded being is shown by inspection of our engines' preconditions. What licenses the inference is its negative use: the audit removes candidate routes to cruelty and installs no positive structure in their place. The licence has its cost in the residue: a being might harbour engines answering to nothing in human motivational architecture, and against these the audit is silent. The residue of this section widens by that possibility; the posit absorbs it.

5.3 The question of grounds

One question remains for the proposal; it should be asked here rather than left to the reader: is there any consideration in favour of the companion God's existence, or is coherence the whole case? The chapter's discipline requires an explicit answer, because §4.2 conceded that the companion God is shown coherent rather than produced by the argument. The hiddenness discussion (§6.6) will ask the reader to weigh that problem against the considerations in the proposal's favour. Those considerations must therefore be named and costed, or the weighing has one empty pan.

Three candidates exist, and each is modest. The first is religious experience. Chapter 14 gives experiences of presence, forgiveness, and encounter a standing the debunking tradition denies them: they are real evaluative events — actual registrations by a real modality — rather than noise to be explained away. Chapter 20 (§7.7) fixed the ceiling: a feeling of divine forgiveness is evidence of the feeling and of its transformative reality; it does not establish its theological interpretation. Within that ceiling a residue of weight remains: a proposal on which

such experiences could be registrations of an actual evaluative engagement — a maximal consciousness responding to the experiencer (§A.3) — accommodates their phenomenology more directly than a proposal on which every such experience is misattributed.

The second is the phenomenology of worship. Wolterstorff (2015) describes liturgy as mutual address, and §A.3 argued that the companion God makes the description coherent where the classical God cannot: an impassible being cannot be addressed in any sense that reaches it. A practice whose internal structure is dialogical fits a theology on which there is a respondent. The third is explanatory fit: the companion God accommodates what believers report valuing most — relationship, presence, being known — without the six incoherences the critique removed. A believer weighing conceptions of God against her own practice therefore has a reason internal to that practice to prefer this one.

The costing is strict, and it doubles as the statement of audience. None of the three is evidence that the companion God exists: experiences underdetermine their objects, a practice's internal structure does not populate the world with its addressee, and explanatory fit selects among conceptions for someone already conceiving. They are considerations of fit, and fit moves only those with something to fit — which fixes who the proposal is for. To the reader antecedently inclined toward theism, the chapter offers a conception on which her experience, her worship, and her sense of relationship are coherent without the claims the critique dismantled, together with the three considerations above, costed as stated. To the reader without that inclination, the chapter offers coherence and nothing further, and does not pretend the considerations should move her. The proposal's register — proposed, the ladder's weakest tier — already encodes this, and §6.6's weighing should be read with these pans loaded just so.

6 Objections and Replies

6.1 The emptiness objection

Objection. A God who did not create the universe, who does not judge by creed, who does not command, and who does not punish is too diminished to be God. The chapter presents a philosophy of empathic engagement dressed in theological vocabulary.

Reply. The objection assumes that the six removed claims are constitutive of theism — an assumption that is itself a product of the conflation of the two infinities (§3). Once the conflation is resolved, a God of infinite diversity — endlessly rich in internal structure, maximally conscious, maximally loving — is recognisably theistic and the appropriate object of reverence: a being whose diminishment is the shedding of incoherence, not the loss of divinity. Mesle (1993) develops the pastoral dimension: a God who participates in events alongside the sufferer rather than controlling them provides a more adequate response to suffering than one who permitted or planned it. The companion God satisfies the requirement — he does not permit suffering, because he did not design the world in which it occurs, and his grief, if the concept is coherent, is real: a being who registers suffering through his own modality, not a designer observing the consequences of his plan.

6.2 The presumption objection

Objection. The chapter presumes to tell believers what their faith should be. Faith is a deeply personal commitment, and a philosophical framework has no standing to reconstruct it.

Reply. The chapter does not prescribe; it derives consequences, and whether a believer revises her commitments in light of them is entirely her decision. The relationship between the chapter and the believer is one of deep persuasion, not enforcement — its authority, like the companion God's, is the authority of a perspective that must earn assent through its quality.

6.3 The coherence objection

Objection. A non-creator God is merely a very powerful being within existence — and calling such a being "God" stretches the concept beyond recognition.

Reply. The objection identifies a real tension. The concept of God in the dominant traditions is linked to creation. The argument is that the link to origination is incoherent, and that the link to sustaining dependence survives only in the first of §4.1's two figures, which the chapter has set aside for stated reasons. The question is whether the concept can survive the severing. The companion God differs from a powerful alien in one decisive respect. An alien is a being whose registration of humanity is partial in principle, bounded by position, interest, and the finite reach of a bounded field, and whose engagement is contingent on the interests a bounded position supplies: territory, advantage, curiosity. The companion God is a being whose registration is unbounded in reach, since no instance of suffering or flourishing lies in principle beyond a field that is fed without limit across time (§5.1), and whose care, though posited rather than derived (§5.2), is structurally unlike an interest: it is the direction of a current that has no position to defend, nothing to gain, and no rival engagement to trade off, and so it cannot be alienated by circumstance as an interest can. The distinction between such a being and a powerful alien is not terminological; it is structural — registration unbounded in reach where the alien's is bounded, positionless orientation where the alien's is interested — with the direction of the orientation carried, as §5.2 records, by the proposal itself.

What matters is the relationship: reverence, devotion, and empathic engagement with a being who transcends ordinary human capacities and who responds to suffering and flourishing with care. A relationship with that structure — reverence answering maximal registration, devotion answering positionless care — is religious in the sense the traditions themselves privilege, and the being at its centre is recognisably God.

6.4 The contingency objection

Objection. Section 4.2 acknowledges that the non-creator God may be contingent. But religious devotion seems to require that God's existence is necessary — that the being one devotes oneself to could not fail to exist. A contingent God is a God one might lose.

Reply. The concern is real, but the assumption that devotion requires necessity may itself be a product of the creator theology the analysis has set aside. We devote ourselves to mortal persons — friends, children, partners — whose existence is contingent and whose loss is possible, and the devotion is not diminished by the contingency; it may be deepened by it (Chapter 20, §3.3). Devotion to a God one is fortunate to encounter rather than guaranteed to possess may be a deeper, more truthful form of religious engagement than one grounded in the certainty that the object cannot be lost.

6.5 The inefficacy objection

Objection. A companion God who acts only through deep persuasion is powerless before suffering, and the powerlessness is not relieved by his perfections but sharpened by them. A being of infinite diversity is maximally knowing: he sees every affliction in full and knows, for each agent, what would move them. Grant him too the power such a nature seems to imply; he could not merely witness suffering but heal it — and the healing would be the most persuasive act of all. A God who could see, relieve, and thereby persuade, yet accompanies the sufferer without delivering her and opposes evil without defeating it, is either impotent or indifferent. Presence is cold comfort, and a divinity that changes nothing is not worth the name.

Reply. This is the gravest objection the companion God faces. The chapter concedes its starting point without qualification: this God neither coerces nor prevents, guarantees no outcome, and offers no theodicy of rescue. What the objection misreads is the passage from that concession to impotence or indifference, and the misreading turns on conflating two perfections the analysis keeps apart.

Omniscience is knowledge, not power over nature. It does make the companion God the optimal persuader: he never misjudges what, if anything, would move an agent. But persuasion terminates in the other's uptake, and knowing the perspective that would move someone is not that person's being moved. To close the gap — to supply the uptake — would be to cause the evaluative change directly, which is the manufactured assent the analysis has already refused (Chapters 16, 21): it would yield compliance but no converted response. He can perfect the offer; the response is constitutively the other's to make. On the book's own conception of a God who grows by accumulating experience (§A.6), an undetermined future choice is not yet there to be foreknown, so even his omniscience does not reach across it. And on the determined branch — the book decides between the branches nowhere (Chapter 18) — foreknowledge of the uptake is not production of it: knowing what will move her is still not her being moved, so the constitutive gap stands on either branch.

The limits of power. Omnipotence is not his to grant. The being who could heal a body or still a catastrophe at will is the maker and master of the causal order — the sustainer of §4.1, whose power over what depends on him returns the problem of evil at full strength, while Chapters 9–12 showed the simple, originating version of that being to be incoherent — and §4.1 and §5.1 build the companion God expressly as a non-sustainer whose sole mode of action is the offer of evaluative perspectives. To equip him with the power to heal is to exchange him for the other survivor and to take on that survivor's cost; the companion God gains nothing by it. And the trade is not incidental: the companion God is innocent of the world's suffering because he cannot prevent it. Restore the power to heal and the problem of evil returns at full strength — an omnipotent healer who leaves a single affliction standing is answerable for it. One may have a God who could guarantee rescue or a God not answerable for its absence, but not both; the position secures its exculpation by its inability. That is a feature of it, not a flaw in it.

The asymmetry — that he reaches wills but not nature — is therefore principled rather than stipulated. His only power is the kind that can be freely declined. An offered perspective is a contribution to

the field a free mind weighs; it can be refused; the healing of a body is a non-declinable override of the causal order, which no one declines. The first is participation in the evaluative life of minds, exercised supremely by a maximal consciousness; the second is the authorship-grade control he does not possess. The same autonomy-preserving structure governs the ethics throughout (Chapter 20): his influence is always the kind that leaves the other free, and for that reason not the kind that secures outcomes.

Nor, were healing available, would it be the persuasive instrument the objection imagines. Relief offered as a reason to accept God's perspective moves the will by power, gratitude, and self-interest rather than by engagement with the relevant evaluative differences — the manipulation the framework excludes (Chapter 21, §7.6), producing allegiance rather than transformation. The narrow exception is the relief of an agony that is itself blocking a sufferer's capacity to engage at all: this restores the conditions for persuasion rather than standing in for it, and it too presupposes a power the companion God lacks. The corollary belongs on the residue list in its own right: a God who cannot relieve even persuasion-blocking agony cannot perform triage of any kind — the most modest intervention the pastoral imagination asks for is already beyond him — and the proposal owns that cost rather than softening it.

The residue owned. What remains is the residue, which the chapter does not disguise: suffering with no will to address, in the natural order he did not design; agents whom no true perspective moves, of whom the coherent outlier of Chapter 14 (§6.4) is the limiting case; and the freedom of uptake that no persuasion, however perfectly informed, can close. The companion God shares this condition rather than standing above it — the contingency conceded in §4.2 is felt here at its sharpest. The chapter does not promise that suffering will end. It claims something more candid and more demanding: that the divine relation to suffering is the one the analysis vindicates as right — to be present to it as its fullest witness, and to oppose it by the only means that transform rather than merely suppress. A God who promised more would have to be the maker and master the preceding chapters have shown such a being cannot coherently be.

6.6 The hiddenness objection

Objection. A companion God whose sole mode of action is the offer of evaluative perspectives faces the problem of divine hiddenness in a sharper form than classical theism does. Schellenberg (1993, 2015) argues that a perfectly loving God would ensure that no one capable of relationship with him remained a non-resistant non-believer, since being open to the relationship is a good such a God would want available to all who do not refuse it. The existence of non-resistant non-belief is therefore evidence against such a God. For the companion God the argument bites harder. His whole relation to persons is persuasive rather than coercive, and persuasion requires that the one to be persuaded encounter the perspective offered. A companion who lures but never appears has failed at the one thing he is said to do. Worse, the framework itself concedes that his existence is an open question (§4.2); a God whose reality cannot be told apart from his absence is just the God the hiddenness argument predicts we should not find if he were there and loving.

Reply. The objection is powerful, and the chapter does not claim to dissolve it. It claims that the companion God is better placed under it than the classical God, and that the residue it leaves is one the framework already owns. Three points locate the position.

First, the sharpest form of Schellenberg's argument is aimed at a God who both loves perfectly and possesses the power to make himself evident to every non-resistant mind at will. That second attribute is the omnipotence the companion God lacks by construction (§4.1, §6.5), and which the sustainer carries at the cost §4.1 records. The companion God does not command the causal order, so he cannot stage the self-authenticating disclosure the argument imagines a loving God would arrange. Hiddenness for such a being is a limit of his mode of being rather than a withheld good — the same limit that makes him innocent of the world's suffering. The argument's premise that a loving God "could" prevent non-resistant non-belief is, for this God, false, and where it is false the inference from love to universal availability does not run. Rea (2018) replies to Schellenberg that divine love need not take the form the argument assumes, since a personal God may have

reasons of his own for the manner of his presence; the companion's position asks less than Rea's, because it needs no such reasons, the manner being fixed by what the companion can do.

Second, the analysis reinterprets what "belief" the relationship requires, and the reinterpretation matters. On the account of faith developed in §A.4, the relationship to the companion God is constituted not by assent to the proposition that he exists but by empathic orientation — the modality's engagement with the suffering and flourishing of others, which §A.4 identifies as the moral content of faith.

A person so oriented is, on the chapter's own terms, already in the relationship the companion God offers, whether or not she has formed the belief that he exists, while a person who assents to his existence and is empathically inert is not. This does not answer Schellenberg on his own ground, since he is concerned with conscious relationship and the reframing shifts what relationship consists in. But it drains the argument of much of its force in the companion case, because the good whose distribution the argument tracks — propositional acquaintance with God — is not the good the companion theology treats as central.

Schellenberg's natural rejoinder should be completed rather than left implied: a loving companion would surely want to be known as one, since wanting the relationship recognised is part of wanting the relationship. The reply is the powers concession already on the table. Wanting is not effecting, so a being who persuades without coercing, and reaches only those whom true perspectives can move, may want a recognition he cannot bring about, for the reasons §6.5 sets out. The exchange ends there, with the want conceded and the ability denied — a completion, on Schellenberg's own terms, of what hiddenness can and cannot show against a companion rather than an omnipotent God.

Third, what remains is a real residue, and the chapter marks it rather than disguising it. If the companion God exists, is loving, and desires that persons come to share his orientation, then his non-evidence to some who do not resist it is a cost the position carries. Smaller than the classical God's, because he lacks the power to remove it; not zero,

because a maximally resourceful persuader might still be expected to reach more than are in fact reached.

The plain terminus is the one §4.2 already stated: the companion God's existence is not established by the analysis but shown to be coherent, so the weight of hiddenness is one of the considerations a reader must set against those collected and costed in §5.3. The chapter's claim is comparative and modal rather than evidential: of the gods that remain structurally available after the critique, the companion God is the one for whom hiddenness is a limitation rather than a withholding; that is the most a theology disciplined by the preceding chapters is entitled to say.

6.7 The divine-suffering lineage

Objection. The chapter places itself against process theology while passing over its nearest neighbours: the twentieth-century theologies of divine suffering. Bonhoeffer (1951/1971) writes from prison that "only the suffering God can help"; Moltmann (1972/1974) makes the cross constitutive of the divine life; Jonas (1987), answering Auschwitz, proposes a God who divested power in creating; Oord (2015) argues that uncontrolling love belongs to the divine essence, so that unilateral prevention was never available. Each arrives at a God who accompanies without rescuing, none by way of Chapters 8 or 10. If the destination is already occupied, the derivation earns nothing.

Reply. The convergence is real, is recorded here, while the ground differs, which is where the derivation does its work. For all four the divine figure remains creator or sustainer, so the inability to rescue is kenotic — chosen in the act of creating, or essential to a nature that might have been described otherwise — which leaves the theologian to say why a God holding that office declines the power the office carries. The companion's inability is structural: it follows from inhabiting the field rather than grounding it (§5.1), so no declining is required, none is available.

The lineages also differ on what is owed. A kenotic creator remains answerable for the creating, which is where Jonas locates the cost,

paying it in divine self-limitation; the companion answers for nothing, since it never occupied the position from which prevention would have been possible (§4.1). What the lineage does establish is that the destination is habitable, which the chapter welcomes rather than resists, since convergence by four routes with different starting points is evidence that the position is no artefact of this framework's vocabulary. The cost §5 itemises stands unchanged, with nothing in the convergence upgrading the posits.

7 Conclusion

The chapter has argued three things. Two structurally coherent theisms survive the analysis's critique on the diversity side, and the chapter has developed one of them. The deep source of the incoherence in classical theism is the conflation of the two infinities — the demand that God be both infinitely diverse (omniscient, omnibenevolent, omnipotent) and infinitely equal (absolutely simple, the undifferentiated ground of being). And this survivor is compatible with structural peace, since the God who persuades rather than commands is the theological counterpart of the deep persuasion peace requires.

The bookkeeping of §1 applies at the close as at the opening. The coherence of the companion God was fixed by arguments that owe nothing to the political analysis; the decision to develop him was guided by those results; and the compatibility is a consequence of that selection — a persuasion-only God fits a persuasion-based peace by construction — rather than a convergence the two inquiries reached independently.

The chapter yields three results.

The companion God (proposed). Once the conflation is resolved, the coherent options are Gods of infinite diversity. The proposal develops the companion rather than the sustainer (§4.1): a companion rather than a creator, who lures and does not command, who responds empathically and does not judge orthotically, and whose relationship to humanity is one of shared existence and mutual evaluative engagement.

The ontological status of this God has been addressed directly. He is within the hierarchy of diversity rather than at its foundation; he may be contingent rather than necessary; he exists within time and responds to events. His mindedness is carried as a posit within the proposal (§5.1), since structure fixes what a maximal mind would register, not that a mind is there to register it. His benevolence is carried as a posit within the proposal (§5.2), since structure fixes the capacity of the divine current, not its direction, and what the chapter derives is only that the limitation-based engines of malevolence are idle in such a being.

Faith as evaluative orientation (constructed; Appendix, §§A.1–A.5). Faith is evaluative orientation toward the transcendent: sustained empathic engagement with a domain of experience that exceeds ordinary categories — a shift, in Chapter 15's terms, from derived-value dominance to primary-value engagement. Concrete practices are preserved and structurally grounded: prayer as evaluative reflection, worship as shared orientation, moral education as empathic formation.

Convergence with process theology (constructed; §A.6). The constructive proposal converges with process theology (Whitehead, Hartshorne, Griffin) on the character of the resulting God while arriving by a different route: the present account of difference, the two infinities, and the conditions for democratic peace. The companion God is not a postulate of process metaphysics. He is the God who remains after the conflation of the two infinities has been resolved in favour of infinite diversity.

The constructive proposal does not ask believers to abandon their faith. It asks them to consider whether the structurally incoherent elements of their traditions are essential to what they love about those traditions, or whether what they love, whether community, contemplation, devotion, moral seriousness, or the experience of the sacred, can be preserved, and perhaps deepened, once the incoherent elements are set aside.

This book began with a question: is there anything that cannot coherently be denied? The answer was difference — the one thing whose denial cannot be truly held, since the denying is an occurrence

and an occurrence is a difference. From that foundation the book built a structural account of content, consciousness, value, morality, freedom, purpose, and peace, each part graded for the weight it bears.

Along the way it showed that the grounding terminus is unintelligent, that creation ex nihilo is incoherent, and that the traditional God of classical theism conjoins properties the structure of difference makes incompatible. It also showed that what remains, once the incompatible elements are set aside, is substantial: a God of infinite diversity, a faith of evaluative orientation, a morality of empathic engagement, and a politics of deep persuasion. The chapter's final claim is not that these results are infallible, since Chapter 17 showed why no such claim can be made, but that they are the best the analysis can offer, subject to revision through the very process of deep persuasion the book has identified as the path to peace.

Grades and Dependencies

1. Posits. The companion God is *proposed* — the ladder's weakest tier — with three posits: mindedness (§5.1: structure fixes what a maximal mind would register, not that one exists; the posit is of a temporal mind, meeting Chapter 13's interface condition diachronically), benevolence (§5.2: capacity derived, direction posited; only the limitation-based malevolence engines are shown idle), and the §5.3 considerations, fit not evidence. Faith as orientation and the process convergence are *constructed*. Order of argument (§1): Chapters 8–12 drew the coherence space; two figures survive on the diversity side (§4.1), a result established by inheritance from Chapters 4–11, and the politics selected between them. The Premise Register carries the posits and the residue.

2. Owned residue. Suffering with no will to address; agents no true perspective moves; uptake no persuasion can close; no triage — even persuasion-blocking agony is beyond him (§6.5). Hiddenness weighs non-zero against §5.3 (§6.6). The §5.2 engine audit is a projection-type inference, licensed only negatively; engines alien to human architecture escape it; the posit absorbs it. The exculpation is secured by the inability (§6.5).

Appendix: Faith Without Obedience

This appendix holds what the companion God makes of faith, worship, and practice, together with the lineage the proposal converges with: consequences of the proposal and no part of its derivation, so that declining the companion forfeits them and no earlier result. Faith is reinterpreted as evaluative orientation (§A.1), the community, ritual, and contemplation it sustains are placed (§A.2), the worship question is answered (§A.3), the reframing of faith without obedience is drawn out into practice (§§A.4–A.5), the process-theology engagement is recorded (§A.6), and the two objections aimed at these consequences are met (§§A.7–A.8).

A.1 Faith as evaluative orientation

Faith, on the analysis's account, need not be understood as belief in a set of propositions — that God exists, that God created the world, that scripture is infallible. It can be understood instead as an evaluative orientation: a sustained engagement of the evaluative modality with something that exceeds ordinary experience. A person of faith is a person whose evaluative modality is oriented toward a domain of experience that she identifies as transcendent.

This is faith as empathic orientation rather than orthotic compliance. The person of faith is oriented — emotionally, evaluatively, with the full engagement of her evaluative modality — toward something she experiences as sacred. The orientation is real, since it involves real evaluative differences detected by a real modality. What the analysis removes is the claim that this orientation must be directed toward a specific doctrinal content. What it preserves is the orientation itself.

In the terms of Chapter 15's value taxonomy, faith-as-evaluative-orientation is constituted primarily by primary value, the evaluative modality's direct response to the transcendent, rather than by derived value, the authority of the tradition, or consequential value, the reward of salvation. The shift from orthotic faith to evaluative faith is therefore a shift from derived-value dominance to primary-value engagement. This is why mystical experience feels different from doctrinal

obedience: the mystic's engagement is constituted primarily by primary value, the obedient believer's primarily by derived value.

This understanding converges with the mystical traditions: the Sufi dhawq, direct tasting of the divine; the Christian contemplative union; the kabbalist's devekut, cleaving to God through direct experience. All of these privilege the modality's direct engagement over orthotic conformity, which is why the analysis explains they have consistently felt different from their doctrinal counterparts: they engage a different evaluative register.

A.2 Religious community, ritual, and contemplation

The preceding chapters established that purpose is constituted by the evaluative modality's engagement with real differences (Chapter 19), and that religious practices work because they engage the evaluative modality. A communal act of worship can therefore produce real empathic engagement: participants share an evaluative orientation, respond emotionally to shared symbols and narratives, and experience a form of collective purpose that individual practice may not achieve. Contemplative practice (meditation, prayer, silent reflection) can heighten the evaluative modality's sensitivity, making the practitioner more attuned to evaluative differences she might otherwise overlook.

These practices do not require the six claims the analysis removes. A community can worship a God of infinite diversity. Contemplation can be directed toward the transcendent without presupposing scriptural infallibility. Moral education can cultivate empathic engagement without grounding it in creed-based judgement. The institutional infrastructure of religious life is preserved. What is removed is the claim that the infrastructure's value derives from a divine source and not from its capacity to engage the evaluative modality with real evaluative differences.

A.3 The worship question

If God is not the creator, and not ontologically superior in the foundational sense — grounded by the same terminus as every other

being rather than grounding them — what grounds the attitude of worship, prayer, and devotion?

The answer draws on the evaluative modality. Worship is constituted not by the worshipper's subordination to a superior being but by her evaluative engagement with something that exceeds her ordinary categories of experience. A person who stands before a mountain and feels awe is not subordinating herself to it; she is responding to something that exceeds her scale, the awe constituted by the modality's response to the mountain's vastness, permanence, and indifference. The response is real, since the mountain does not need to be her creator for the awe to be real.

Wolterstorff (2015) develops an account of liturgical action as a form of mutual address between God and the worshipping community: more than merely human speech directed at God, it is a dialogue in which both parties are affected. The companion God makes Wolterstorff's dialogical account more coherent than classical theism does, because a classical God, impassible and unchanging and already possessing all knowledge, cannot be affected by the dialogue. A companion God can be, since his evaluative modality is engaged by the community's expression, so the dialogue is structurally bidirectional.

Worship and inexhaustibility. Worship of the companion God has the same evaluative structure with immeasurably greater depth. A being of infinite diversity, infinitely rich in internal structure, maximally conscious and maximally evaluative, presents the evaluative modality with a field of differences that exceeds any finite engagement and cannot be exhausted by any engagement, however prolonged.

The reason is the saturation analysis. Saturation (Chapter 20, §3.2) is the running-out of a field's novel differences, and an infinitely diverse field has no exhaustion point, so this is the one form of unending engagement the saturation analysis leaves standing — the property Chapter 20 (§6) uses for its universalism. The awe this engagement produces is the awe of encountering a depth of evaluative reality that transforms the worshipper's own evaluative perspective.

This account must be distinguished from mere aesthetic appreciation. Aesthetic appreciation, as the value taxonomy of Chapter 15 makes clear, is a form of evaluative engagement, while worship of the companion God involves a further feature: reciprocity. The worshipper's evaluative engagement is met by the companion God's evaluative engagement with the worshipper. Aesthetic appreciation of a sunset is unidirectional, since the sunset does not evaluatively register the observer.

Prayer, on this account, is an act of evaluative reflection — an orientation of attention toward what matters. To pray is to direct the evaluative modality's engagement toward the companion God's perspective: to try to see as he sees, to be moved as he is moved, to care as he cares. This is prayer as evaluative cultivation. Petition has dropped away, at a large cost: a God who cannot act on the causal order cannot answer a request to do so, so the prayer of the frightened and the bereaved — the prayer most people most often pray — has no addressee on this proposal, and the residue list of §6.5 includes it.

A.4 The structural reframing

If morality is constitutively empathic (Chapter 16), then the moral content of faith is empathic engagement, not obedience: a faithful person is one whose evaluative modality is deeply and sustainedly engaged with the suffering and flourishing of others, her empathic orientation enriched and sustained by the relationship to God. This reverses the theocratic inversion in which devotion to God takes priority over love for persons and creed outranks character. The companion God does not demand that his followers prioritise allegiance to him over care for others. His own engagement is directed toward the flourishing of the beings alongside whom he exists — the orientation §5.2 records as the proposal's posit — and to be devoted to such a God is to share his orientation: to be moved, as he is moved, by the suffering and flourishing of others.

A.5 Concrete practices

What would a religious practice built on evaluative orientation instead of orthotic compliance look like concretely?

Prayer becomes evaluative reflection — an act of orienting the modality toward what matters. The contemplative traditions already approximate this: centering prayer in Christianity, dhikr in Islam, hitbodedut in Judaism, and zazen in Zen Buddhism all involve the sustained direction of attention toward a domain of experience that exceeds ordinary engagement, and the practices work because they heighten the modality's sensitivity — training it to detect what it might otherwise miss.

Worship becomes shared evaluative engagement — a community orienting together toward the transcendent. The communal dimension matters: shared engagement produces empathic resonance, which deepens the engagement beyond what individual practice achieves; a congregation that worships together is cultivating a shared evaluative orientation that shapes the community's moral life.

Moral education becomes empathic formation. The stories, parables, and exemplary lives that populate religious traditions are effective moral tools because they engage the modality empathically — presenting situations of suffering and flourishing that train its response (Chapter 16).

Faith without obedience is not faith without discipline. The cultivation of empathic sensitivity requires practice, effort, and sustained attention — just as the cultivation of any modality does. A musician's sensitivity to auditory differences requires years of training; a person of faith's sensitivity to evaluative differences requires years of moral and contemplative practice. What changes is the locus of discipline: it is the internal work of deepening the evaluative modality's engagement with the world.

A.6 Engagement with process theology

This conception converges with process theology's vision of God as a persuasive rather than coercive presence. But the present account

arrives at the convergence by a different route: through the account of difference, the two infinities, the evaluative modality, and the conditions for peace, rather than through a metaphysics of process. Open theism reached a temporal, responsive, non-simple God without Chapters 1–8; what those chapters add is the reason the simple and timeless God is unavailable, and the interface condition (§5.1) that makes the surviving God temporal by necessity.

Whitehead (1929) distinguishes two natures of God. The primordial nature is God's eternal conceptual vision of all possibilities; the consequent nature is God's actual engagement with the world, affected by what happens in it. The companion God corresponds most closely to the consequent nature: a being engaged with the world, affected by its suffering and flourishing, responding through persuasion rather than coercion.

The primordial nature corresponds to what the analysis would describe as the evaluative modality operating at infinite diversity: a modality that detects all possible evaluative differences, the most comprehensive evaluative engagement conceivable. The convergence is real while the grounding differs. Whitehead derives God's natures from a metaphysics of process and creativity; the analysis derives the companion God from the account of difference and the resolution of the two-infinities conflation.

Hartshorne (1948) develops dipolar theism: God is both abstract, unchanging in his essential character, and concrete, changing in his actual engagement with the world. The companion God is dipolar in a related sense. He is abstractly characterised by infinite diversity, since his internal complexity does not diminish; concretely, he is engaged with the world in ways that change as the world changes. The evaluative modality's engagement with the world is not static, because it responds to new differences as they emerge. That responsiveness is the temporal interface §5.1 requires of any mind, the maximal one included. A companion God who is infinitely diverse and also temporally engaged is therefore a God who grows, in the sense of accumulating experiences and deepening his engagements rather than of becoming more divine.

One divergence of content belongs on the ledger beside necessity and grounding, because the convergence framing would otherwise paper it over: panexperientialism. Whiteheadian process metaphysics builds the world from occasions of experience all the way down. The framework excludes experience below the categorical transition for composed minds (Chapters 6 and 13), with only the constitutive-micropsychist variant surviving, at the scope Chapter 6 concedes. The companion God therefore inhabits a world most of which, on this account, experiences nothing — a flat contradiction with the process picture of nature. It is a divergence a process theologian should weigh alongside the two the chapter already records.

Griffin (2001) develops what he calls "reenchantment without supernaturalism", a theology that preserves the sense of the sacred without invoking supernatural intervention. The constructive theology developed here is the closest existing parallel to Griffin's project, a convergence that deserves explicit acknowledgement. What the analysis adds to Griffin is a foundation. The analysis of difference, the two infinities, the evaluative modality, and the conditions for democratic peace provide a systematic basis for the companion theology that Griffin develops on more intuitive grounds.

A.7 The pastoral concern

The objection is that the chapter's analysis is intellectually coherent but pastorally insensitive. For many believers, releasing the beliefs under critique — creation, infallibility, creed-based salvation — involves profound grief. The argument does not adequately address this grief.

The concern is real and the chapter acknowledges it without reservation. The transition is not merely intellectual: the believer is losing secondary objects — the derived value of traditional associations with a creator God, an infallible scripture, a salvific creed — that have populated her hierarchy for a lifetime, and the grief is real because the loss of derived value is a real evaluative diminishment. But the loss is not total. The primary value of her engagement — her direct response to the transcendent, her empathic engagement with the community, her

contemplative orientation — survives the transition, and the tradition that emerges may be deeper, not shallower, than the one it replaces, because its engagement is grounded in primary value rather than in derived value resting on structurally incoherent foundations.

A.8 The process theology divergence

The objection is that the chapter claims convergence with process theology but does not address the points of divergence: does the analysis accept Hartshorne's dipolar theism in full? Does it accept Whitehead's concept of God as the "lure" of creative transformation?

The convergence is real but partial, and the divergence can be stated cleanly. Hartshorne's God is metaphysically necessary — his abstract nature necessary, his concrete states contingent — where the companion God may be contingent in both respects (§4.2). Hartshorne's God is the ground of the metaphysical categories (creativity, eternal objects); the chapter's God is within a hierarchy whose ground is unintelligent equality. Hartshorne's God is dipolar by metaphysical argument; the chapter's is a being of infinite diversity by structural derivation from the two-infinities distinction, arrived at through the analysis of difference rather than through Whitehead's metaphysics of creativity and eternal objects. The two routes converge on a remarkably similar destination — a temporal, responsive, persuasive God — but the analyses are independently motivated and make different commitments about necessity, grounding, and the categories.

References

Bonhoeffer, D. (1951/1971). *Letters and papers from prison* (E. Bethge, Ed.). SCM Press.

Buckareff, A. A., & Nagasawa, Y. (Eds.). (2016). *Alternative concepts of God: Essays on the metaphysics of the divine*. Oxford University Press.

Griffin, D. R. (2001). *Reenchantment without supernaturalism: A process philosophy of religion*. Cornell University Press.

Hartshorne, C. (1948). *The divine relativity: A social conception of God*. Yale University Press.

Hasker, W. (2004). *Providence, evil and the openness of God*. Routledge.

Jonas, H. (1987). The concept of God after Auschwitz: A Jewish voice. *The Journal of Religion, 67*(1), 1–13.
Keller, J. A. (2007). *Problems of evil and the power of God*. Ashgate.
Leftow, B. (2012). *God and necessity*. Oxford University Press.
Mesle, C. R. (1993). *Process theology: A basic introduction*. Chalice Press.
Moltmann, J. (1972/1974). *The crucified God: The cross of Christ as the foundation and criticism of Christian theology* (R. A. Wilson & J. Bowden, Trans.). SCM Press.
Oord, T. J. (2015). *The uncontrolling love of God: An open and relational account of providence*. IVP Academic.
Pruss, A. R. (2006). *The principle of sufficient reason: A reassessment*. Cambridge University Press.
Rea, M. C. (2018). *The hiddenness of God*. Oxford University Press.
Schellenberg, J. L. (1993). *Divine hiddenness and human reason*. Cornell University Press.
Schellenberg, J. L. (2015). *The hiddenness argument: Philosophy's new challenge to belief in God*. Oxford University Press.
Ward, K. (2007). *Divine action: Examining God's role in an open and emergent universe*. Templeton Foundation Press.
Whitehead, A. N. (1929). *Process and reality*. Macmillan.
Wolterstorff, N. (2015). *The God we worship: An exploration of liturgical theology*. Eerdmans.

General Conclusion

1 The Shape of the Argument, Revisited

The introduction gave the argument the shape of an arch: difference as the bedrock, a footing, two sides each rising as a pier and turning as a curve, and peace as the keystone where they meet in Chapter 24 (General Introduction, §3). Twenty-four chapters later, this section says where each part stands, at the grade the Register records. The table repeats the Introduction's, from bedrock to keystone, with the verdict on every claim; the Premise Register carries the verdict on every supporting result beneath the same claims.

Part of the arch	Claim (chapter)	Verdict
Bedrock	At least one difference obtains: the one thing that cannot be denied. (1)	Established; the premises self-vindicating
The footing (Chapters 2–7).		
Footing: ontological (2–6)	Content requires contrast, and the diff is the unit of content. (2)	Established by forced choice; the critic's shortest route in (§2)
	The diversity/equality partition is exhaustive. (3)	Established by the instantiation thesis and the tests
	Equality is the foreclosing terminus of grounding. (4)	Established, conditional on the termination branch; limitarianism holds on the constitutive order alone (§3)

	Time terminates at the same boundary, and the two hierarchies converge. (5)	Established on the termination branch, with a comparability assumption named
	The foreclosing terminus is unintelligent. (6)	Established; the zero-complexity constraint displays a presupposition floor
Footing: epistemo-logical (7)	Meaning and difference are one closed grid. (7)	Established at the tier of Chapter 2's thesis (§2)
The theological side: no mind stands above the hierarchy of difference (Chapters 8–12, 17, 20, 23).		
Pier: core theology (8–12)	"Infinite" names two opposite conditions, and the terminus has three names. (8)	Established; its diagnostic use conditional on the richness premise (§2)
	Creation is dependence, and the sustainer survives the transition picture. (9)	Established against the transition picture; the mature doctrine survives renamed
	Ground and mind come apart: the simple God is incoherent, the determinate God an inhabitant. (10)	Established on the identity reading; the containment fork costed (§2)
	The kalām reaches no necessary being. (11)	Dialectical: the burden relocated, the thesis unrefuted
	Negative theology names a limit of thought and misreads it as a being. (12)	Established; the eminent reading granted

Curve: applied theology (17, 20, 23)	Perfection does not entail infallibility; dogma has a four-stage anatomy with a refutation condition. (17)	Classification established; anatomy constructed, its refutation condition stated (Chapter 17, §6)
	Desert fails by two routes, so restoration replaces retribution. (20)	Constructed; each route at its chapter's tier; the terminus proposed
	Creed-based judgement is the conflation projected onto earthly standing: the divine wall. (23)	Constructed; staked on the doctrinal gradient (Chapter 23, §3.5)
The metaethical side: no moral order stands outside the parties (Chapters 13–16, 18–19, 21–22).		
Pier: core metaethics (13–16)	The mental has two axes, content and orientation, and value lies on the second. (13)	Partition derived; the orientation axis constructed, the joint §2 names
	Value is constituted by the evaluative modality's response and anchored in the affected party; the moral is the empathic evaluation of chosen behaviour. (14)	Constructed; one contested premise, that favouring is grip; the counterfactual residue carried as a limit
	Total value composes primary and added value; across evaluators a trade-off has no fact beyond convergence under mutual registration. (15)	Constructed; the convergence claim is the thinnest empirical bet
	The major ethical traditions mistook conformity to a standard for empathic evaluation. (16)	Constructed; a misattribution of source, staked on Chapter 14 for the rest

Curve: applied ethics (18, 19, 21, 22)	Authorship is bounded: freedom is real from the first-person present and determined from the third-person retrospect. (18)	Constructed; a development within hard incompatibilism
	A purpose is an object with a sense; it requires no endorsement from outside the field. (19)	Constructed; the normative question declined
	Structural peace comes only through deep persuasion, the transformation of the preferences that generated the conflict. (21)	Constructed; analytic for the necessity of deep persuasion, empirical for its occurrence
	The conditions for deep persuasion are analytic and its institutions means; relativism and absolutism share the authority-dependence that disables lawmaking. (22)	Analytic for the conditions; the diagnosis explains a pattern it does not derive (Ch. 22, §7.4)
Keystone: peace (24)	The God that survives persuades and does not command: the companion. (24)	Proposed; three posits, costed
	Peace, where the two sides meet. (21–22, 24)	Analytic where it claims necessity, conditional on the aim; proposed where it posits

The bedrock. Established, with its premises at the self-vindicating tier: the denial of difference instantiates what it denies, and the ontological step from representational to real difference is secured by the occurrence premise and the occupancy test (Chapter 1). The fourth category itself is established, since the mechanism that individuates it runs through Chapter 2's thesis (§2).

The footing. Established. The partition is formally exhaustive and made substantive by the constitutive instantiation thesis; the termini follow, equality as the foreclosing terminus of grounding and the

boundary of time, unintelligent because no cognitive property can characterise a state with no distinctions (Chapters 3–6); the grid of meaning is closed at the tier of Chapter 2's thesis (Chapter 7). The footing owes four things: the termination branch is a conditional, the count-is-one result rests on a comparability assumption, the two-names result rests on the premise that a lack confers no internal articulation (Chapter 3, §3.5), and the something-from-nothing question keeps one brute fact, the obtaining of differentiation (Chapters 8 and 9).

The theological side. Its pier is established for the identity reading of simplicity and costed for the other. Read as the absence of internal distinction, the simple and maximally intelligent God conjoins incompatible properties; read as maximal determinacy, the God is a structured inhabitant, and the containment fork states what that reading costs the doctrine of virtual containment (Chapter 10, §3.2; §2). The intra-attribute instrument is the form of the case against simplicity, where Plantinga and Mullins made the substance. The kalām reaches no necessary being, a dialectical result whose burden is relocated while the thesis stands unrefuted (Chapter 11). Negative theology reads a limit as a being (Chapter 12). Creation as a transition dissolves while dependence without origination survives under a new name (Chapter 9). Its curve (Chapters 17, 20, 23) is established where it classifies and constructed where it applies: infallibility falls to perfection, desert fails by two routes each at its chapter's tier, and the divine wall is staked on the doctrinal gradient Chapter 23 names (§3.5).

The metaethical side. Its pier is constructed at the tier of the response-constitutive account (Chapter 14), and everything applied inherits that tier: the vacancy result empties the clause every rival needed, and the affected party's registration anchors the moral fact against relativism (§§6.4–6.5). The vacancy, containment, and quarantine results leave a critic two tolled exits and one refusal, the premise that favouring is grip, carried by name in the Register. The conflation the major traditions share is diagnosed at that cost (Chapter 16); one thing the account leaves open is inherited by every applied chapter: registrations do not sum, so a trade-off between parties has no fact beyond what mutually informed fields settle on (Chapter 15). Its curve (Chapters 18, 19, 21, 22) is constructed where it claims results and analytic where it claims

necessity: bounded authorship is a development within hard incompatibilism, a purpose is an object with a sense that needs no endorsement from outside the field, structural peace comes only through deep persuasion, and the conditions for it follow from what deep persuasion is.

The keystone. The political crown of the metaethical curve held as far as it claimed: the three conditions are analytic and conditional on the aim, the institutions that secure them are the best available means, and the diagnosis of bipolar authoritarianism explains an oscillation it does not derive (Chapters 21–22). The keystone itself (Chapter 24) is established in what it eliminates and proposed in what it keeps: the two survivors are established, the companion is a proposal with its mindedness and benevolence posited, and faith as evaluative orientation is constructed. Its weight rests on both sides, the theological pier for what survives and the metaethical curve for the persuasion the companion practises.

The joints of the arch are where it is tested, and they held unevenly. Bedrock to footing held at the established tier. Footing to the metaethical pier is the first constructed joint, since the premise that orientation is a determination's only further division is constructed and used (§2); pier to curve held at the constructed tier, with the trade-off residue passed along; the curve's political crown held on the anchor and the field. Footing to the theological pier held on the unintelligence of the terminus; pier to curve held on the perfection/infallibility distinction and on the wall cleared, staked on the doctrinal gradient. Both curves to the keystone held on the two survivors and on deep persuasion, which the companion practises. So the arch stands as the Introduction said it would: established below, constructed on its metaethical side, and proposed at the keystone, which posits what it cannot derive.

One claim is entered in the table as dialectical, so that the entry is not mistaken for a ranking. The kalām engagement of Chapter 11 — the demonstration that the argument's necessity ascription stands or falls with the impossibility of undifferentiated being, and that this impossibility is a heavyweight modal thesis owed and undischarged — is, in the author's judgement, the book's sharpest single exchange; its

victory is dialectical in the sense Chapter 1 (§2.4) distinguishes, since the burden is relocated onto the proponent and the thesis itself stands unrefuted, which is why the table marks it as it does.

2 Where the Argument Can Fail

The faults a hostile reviewer would find run through the arch part by part; each is a joint the Register names: at the bedrock and the footing, the reach of the self-vindicating tier and the determination dilemma everything downstream inherits; on the metaethical side, the response-constitutive account and the premise that coupling is binary; on the theological side, the containment fork and the residues the companion carries; at the political crown, the conditionality of the political derivation; at the keystone, the companion's posits. The first is placed first because every later grade inherits it.

The reach of the self-vindicating tier (Chapters 1, 2, 7). The framework's foundation is two premises and the one result built from them. The occurrence premise says that an entertaining is an occurrence. The composition premise says that an entertaining of the denial has at least two components. Each is self-vindicating, and holds against any critic whatever. The foundational result, that at least one difference obtains, stands with them at the top tier, since the denial of difference is a real act with two components, and two components of a real act are a real difference.

Everything else stands one tier lower. The vindication law and the exhibitive closure of Chapter 7 run through the structural dependence of content on contrast, which Chapter 2 secures by forced choice, so the inheritance rule places them there, since no mechanism is safer than the premise it travels through.

A reviewer looking for the shortest route into the book should take it here, where the route is marked. Reject the determination dilemma (Chapter 2, §2.4), and the fourth category, the exhibitive closure, the anchoring requirement, and the vacancy result that runs through them come down together. What she cannot reach through that door is the ontological result, which owes Chapter 2 nothing, nor can she reach it

by rejecting the two-names result, which carries only the second of its two routes. The route is set out here so that a reviewer need not find it.

Two further joints belong on the same map, because later chapters use them and neither is named as a gate elsewhere. The first is the containment fork (Chapter 10, §3.2), on which the separation of ground from mind, the richness premise of Chapter 8, and the structural half of Chapter 17's two-infinities argument all rest: its first horn is refused by the Thomist doctrine of virtual containment, and the Register (row 10 §3.2, note w) states the cost of that refusal. The second is the premise that coupling is binary — that a determination's only further division is by felt orientation (Chapter 13, §2.3) — which the Register carries at Chapter 13's constructed tier; what the metaethics draws from Chapter 13 is the determination-or-nothing partition beneath it, together with divisions the trilemma and the grip argument supply for themselves, so the vacancy, parity, and quarantine results are insulated from the premise, while the two-axes architecture as a positive theory is not. A reviewer who wants the load-bearing joints should take these two first.

The companion God (Chapter 24). The constructive theological proposal is structurally coherent, since the resolution of the two-infinities conflation in favour of infinite diversity leaves two conceptions standing, a sustainer and a companion, where the chapter develops the one that is temporal, responsive, and persuasive rather than timeless, impassible, and coercive.

The chapter meets the charge that the framework exempts itself from its own standard: the companion God is contingent where a brute posit would be unexplained, an inhabitant of the diversity hierarchy, bounded by the same equality terminus as everything else and closing no explanatory regress the unintelligent terminus has not already closed.

What remains open is narrower and real: how such a being emerges within the hierarchy, whether it is necessary in any deeper sense, and whether the concept is recognisably theistic. The convergence with process theology is real but partial, and a process theologian would still find the divergences on necessity and grounding underexplored. The selection between the two survivors is made on political grounds and is

stated as such (Chapter 24, §4.1). A reader who prefers the sustainer, with the problem of evil that returns with him, holds a conception the framework leaves coherent, where the reframing around peace adds no exposure beyond that choice. The residual weakness is therefore the ontological one just described.

One element of the residue is sharper than the rest and is named in the chapter itself (§5.2): the direction of the divine current. Structure fixes the companion God's registration while leaving its orientation open, since Chapter 13 makes empathy and cruelty one capacity in two directions and Chapter 14 concedes the coherent outlier. Benevolence is therefore a posit within the proposal, with the limitation-based engines of malevolence shown idle while the pure inversion is not excluded.

The same costing covers mindedness itself (§5.1), since infinite diversity does not entail that the diversity composes a mind, so the proposal posits the instantiation-as-detection and collects Chapter 13's conditionals rather than performing derivations. The pastoral weight of Chapters 20 and 24 rests on those posits, and a hostile reviewer should be directed to them first.

The political derivation of democratic peace (Chapters 21–22). The derivation claims three things at three different strengths.

It claims necessity only for the functional conditions of deep persuasion — expressibility, receptivity, and non-exclusion — and treats free speech, open speech, and democratic equality as the institutions that secure them, offered as the best available means rather than as entailments. Equal deliberative standing is derived from the binary requirement that a perspective can or cannot enter the field, rather than inferred from the graded property of empathic intensity. And democracy is reconstructed as two layers: a necessary deep layer that institutionalises the conditions for persuasion, and an optimal-not-necessary shallow layer that aggregates the unresolved residue by vote.

The schematic treatment of institutions is therefore deliberate rather than a gap, since the institutional layer is instrumental and empirical by design. Nor is the aim of peace asserted, since the chapters are explicitly conditional: if aligned peace is the goal, this is the route. The goal is

itself treated as a target of deep persuasion, so the analysis commands nothing and is performatively consistent with its own rejection of authority-grounded value.

This substantially answers the is/ought and cultural-imperialism worries, since the necessary layer is culture-neutral and a society may satisfy it through institutions other than the Western liberal ones.

What remains is a real but relocated vulnerability. The political results are conditional on the empathic metaethics of Chapters 14–16 at the floor's tier: what they use is the tracking floor, row 14 §3 (floor) of the Register — that empathy registers moral value and that the moral concerns conduct's bearing on those it affects — with the anchor's observer-independence, which a realist about welfare also holds. The floor is what the containment result (Chapter 14, §3.7) leaves the realist, so whoever holds it while refusing constitution refuses an identification and adopts no rival account of value; the quarantine below concerns the rival. The series therefore survives the refusal of constitution and fails only if the floor is refused, or if the premise that favouring is grip is rejected where Chapter 21 (§6.2) uses it; non-exclusion, though principled, is where a determined critic will still press. The second objective in its strong form, that no moral order stands outside the parties, is constitution's; what the political crown uses is the weaker form the floor delivers, that no such order settles a dispute until the parties register it, and persuasion transforms preferences whether or not there is a fact they answer to.

And Chapter 16's conditionality has a fixed shape, given by the quarantine of the field (Chapter 14, §A.4). The empathic identification remains constructed, while the metaethics beneath it has no contentful rival. The applied sequence therefore cannot be declined for an alternative account of value, but only at the chain itself, through the production challenges recorded in Chapter 14's register entry, whose burden lies with the producer.

A hostile reviewer's route into the political chapters accordingly runs through three named gates: exhibit the favouring the trilemma forecloses, exhibit suffering that occurs while nothing is undergone, or

refuse the field/pull model of "must" (Chapter 14, §A.5) and bear the cost of the word.

The partitions, by kind. Six partitions carry the book's closure arguments; they are not one kind of claim, so they are graded apart. The diversity/equality partition is excluded middle and stands ungraded (Chapter 3, §2.2). The three grids are derived from the biconditional (Chapter 7, §3). The determination-or-nothing partition is derived and its orientation cell constructed (Chapter 13, §2.3). The trilemma's division, transferable across registration systems or not, is drawn by logical form, while its identification of the second cell is substantive (Chapter 14, §3.5). The three names of the terminus are a classification argued case by case (Chapter 8, §2). And the three modes of conflict resolution partition a model the book built, on one premise a challenger may refuse — that conduct against a standing preference is sustained only through capacity or calculation (Chapter 21, §3). Where every opponent seems processed, the reason is here: the machinery stops where a partition's premise is refused, with the premise named in each case.

The falsifiability belt, counted. The framework surrounding the partition remains falsifiable in the ordinary way, and the book says where. Chapter 3 (§5) separates a rigged test from an empty one and hands the inventory here, so that it is counted once for the whole book; two candidates must be struck from it before it is offered, since a belt inspected only by the reader who already doubts it is worth nothing.

The partition's own counterexample is the first: a state neither differentiated nor undifferentiated is an empty specification by theorem, and no part of the belt. The zero-complexity constraint states a refutation condition explicitly (Chapter 6, §5) — a credible attribution of consciousness to a system with zero internal state-distinctions — while the same section establishes that no quantitative theory of consciousness could supply one without abandoning state-spaces altogether, which makes the condition a second empty specification. And the equalisation phenomenology could have gone otherwise, since the progressive removal of contrasts might have converged on self-standing minimal qualia and did not, while the constraint is expressly

not rested on it, so its going otherwise would have cost corroboration rather than the result.

What survives is real and is stated at its true width. The consistency model of Chapter 3's appendix could have failed to exist. The obtaining residue (Chapter 5, §2.5) would grow beyond a single featureless fact if the temporal series proved closed below rather than open. And the phosphene prediction of Chapter 13 (§A.7) discriminates the framework's account of consciousness from its nearest neighbours by an observation that could go against it. The test is presented there as well-posed and not yet adjudicated, the stimulation evidence still wanting the attentional controls the decision needs, with the design that would supply them specified in the same section; that is the standing of a live prediction rather than a settled score.

The framework's immunity is therefore located, and the belt around the footing is thinner than a first inventory suggests. Two specifications are empty — the partition's counterexample by theorem, the zero-complexity constraint's refutation condition by the presupposition floor — and what stands in the ordinary line of fire is the consistency model, the obtaining residue, and Chapter 13's prediction with its design. Three exposures are fewer than the framework would wish and more than an unfalsifiable machinery has; the count for the footing is given at three. The applied side's stakes are counted apart, at §4 and in the Register: the doctrinal gradient, the canvassing convergence, the institutional predictions, the clinical divergence of Chapter 14 (§A.6), and the introspection shape of Chapter 18 (§8.9).

3 What the Framework Has Not Achieved

Three limitations deserve explicit acknowledgment; a fourth, of a different kind, stands beside them.

The grounding disjunction. The framework has not established whether the grounding hierarchy terminates. The disjunctive result — either termination in equality or infinite diversity — is real, while the framework does not determine which branch obtains.

The ecumenical structure is a strength, since the results hold on both branches, though it leaves the most fundamental metaphysical question open. And beneath it a second stands recorded where it is costed: whether undifferentiated being is metaphysically possible at all, the commitment Chapter 4 (§4) enters in its own ledger and Chapter 11 (§4.3) weighs, with the burden shown to lie with any party who needs the question closed.

Depth in the subfields. The treatment of individual subfields is necessarily less thorough than a dedicated monograph would provide. The consciousness chapter (13) would benefit from deeper engagement with the neuroscience. The ethics chapters (14–16) would benefit from fuller treatment of the fitting-attitudes debate and the recent experimental work on moral judgement. The political chapters (21–23) would benefit from engagement with specific institutional designs and empirical case studies. These are debts the framework acknowledges.

Adjudication between hierarchies. The framework does not provide a standpoint-independent mechanism for adjudicating between conflicting evaluative hierarchies except through deep persuasion: the process by which one hierarchy is transformed through engagement with another. That process is the mutual registration Chapter 15 (§4.4) names, and its convergence is the only fact a trade-off has.

For critics who demand a neutral moral criterion that can settle disputes without the transformation of preferences, this will be unsatisfying, and the framework's accounting of the demand answers it part by part. The demanded criterion is specified by a clause the framework shows to say nothing (Chapter 14, §3.6). The demand that it bind the unmoved is shown to be a demand nothing anywhere meets (§A.2). The discourse said to presuppose it carries no such commitment (§A.3). And the quarantine of the field (§A.4) leaves the demander no rival position from which the demand can be issued as a theory, while §A.5 states the cost of the premise he would have to refuse to issue it at all.

What remains to the realist is a report: that the vacant clause seems meaningful to him — a seeming the framework predicts, explains as the

projection of §3.5, and cannot make impossible, since arguments operate only on the pulled.

The residue is therefore narrower than a standoff and is recorded at its width: one account, two tolled exits, and one report answerable to no content the framework can locate, with the framework answerable, permanently, for the possibility that the location has failed. A challenger who would widen the residue has the two production challenges, with the burden on the producer, and the two tolled exits, at their stated costs; and recording the residue rather than declaring it closed is itself an application of the book's method.

A fourth limitation is the one the metaethics entails rather than leaves open; it should stand beside the other three at its true size. Registrations do not sum across their bearers (Chapter 14, §6.5), so there is no fact, independent of every weigher, about how one party's agony stands against a million inconveniences.

What the account offers instead is convergence. Within one evaluator, a field is enriched until it tips. Across evaluators, each party takes in the other's registration, so the verdict on a trade-off is whatever such mutually informed fields settle on (Chapter 15, §4.4). Where they settle on the outcome, that is the fact. Where they settle only on how to decide, the procedure is — the convergence one level up on which Chapter 22 builds equal standing. Where they settle on neither, the framework returns no further fact, and Chapter 21's other modes remain.

The convergence is staked empirically, with the book's other exposures. A reader who wants a theory of trade-offs will find here a theory of what trade-offs are made from, of the operation that settles them where anything does, and of the discipline that governs it — and nothing that ranks the outcomes in advance. Every applied result inherits the residue, which is constitution's: refuse Chapter 14 for a realism about welfare, and there is a fact where the account has convergence, so the limitation lifts, while the vacancy result is owed an answer. It is where a critic who grants the metaethics entire will still have most to say.

4 What Would Strengthen It

The faults of §2 have remedies, and most are empirical. The stakes the Register lists are the tests the framework owes: the phosphene prediction for the uniqueness principle, the doctrinal gradient for the divine wall, the canvassing convergence for the metaethical side's thinnest bet, the four institutional predictions of Chapter 22 once they are operationalised. The constructed joints — the orientation axis, the containment fork's first horn — could be argued where they are costed. The directions that follow are the ones the framework itself points to.

A framework that claims to illuminate six domains incurs six debts.

Metaphysics. The grounding limitarianism of Chapter 4 is set against the defence of infinite descent (Chapter 4, §7.1) and should be carried further into the debate about well-foundedness and infinite grounding chains.

The unilateral boundary concept distinguishes a bilateral boundary, where explanation is declined, from a unilateral boundary, where explanation is foreclosed, and defines the latter in the containment order of Chapter 4's Appendix as the non-member bound of a descending chain. It may have applications beyond the grounding hierarchy, wherever a distinction-dependent relation reaches a state without distinctions, which should be explored. The two-names result of Chapter 3 (§3.5), on which much of the theological sequence depends, has begun its dialogue with the metaphysics of absence, where it names the premise it uses, and would benefit from carrying that dialogue into noneist logics.

Philosophy of mind. The categorical transition principle (Chapter 13) generates testable predictions about the structural conditions for consciousness. The falsification conditions — particularly the prediction that Level 4 consciousness cannot obtain without Level 2, and the prediction that experience vanishes categorically rather than gradually when contrast is removed — are the most empirically tractable elements of the framework. The uniqueness principle's three levels, positional, functional, and structural, together with the identity hypothesis that the

equalisation constraint favours, deserve sustained development in dialogue with the neuroscience of consciousness.

Philosophy of religion. The ground/mind conflation (Chapter 10) and the structural response to the kalām (Chapter 11) are the results most likely to attract engagement from analytic philosophers of religion. The companion God (Chapter 24) deserves sustained development in dialogue with process theology, open theism, and the growing literature on non-classical theism. And the Palamite essence/energies distinction, identified in Chapter 10 as the most coherent option available to the theist, merits a dedicated treatment that the chapter's broader scope did not permit.

Ethics. The great moral conflation (Chapter 16) should be deepened into a standalone treatment that develops the response-constitutive account of value (Chapter 14) with the thoroughness the metaethics literature demands. The primary/added value distinction (Chapter 15) has applications to adaptive preferences, the psychology of decision-making, and the philosophy of marketing that the chapter identifies without developing.

The engagement with the fitting-attitudes literature and with the recent empirical work on emotion and evaluative judgement are the most productive directions. The vacancy and containment results (Chapter 14, §§3.6–3.7) now stand as the account's sharpest exposed face, and direct exchange with the robust realism they meet is the test they invite.

Political philosophy. The structural derivation of the conditions for democratic peace should be tested against the political science literature on democratic transitions, deliberative polling, and institutional design. The institutional question — how deep persuasion operates at scale, and what institutional structures best support the transition from managed conflict to real resolution — is the framework's most consequential open problem.

Free will. The perspectival asymmetry and the ethereal mind argument (Chapter 18) should be developed into a standalone treatment that engages the experimental philosophy literature on folk intuitions

about free will and moral responsibility. The distinction between practical and absolute moral responsibility has implications for criminal justice, restorative justice, and the theology of divine punishment that Chapters 18 and 20 identify but that deserve book-length treatment.

5 A Closing Statement

Conflict and disease are the oldest and most expensive of human problems, and the instruments we have brought to conflict are mostly instruments of management: force, which suppresses a disagreement without resolving it, and negotiation, which trades positions while leaving untouched the preferences that produced them.

If the chain in this book holds, resolution has conditions — statable, checkable, and derived from a fact nobody can deny. Peace would then be something a society can be built towards by argument, rather than merely hoped for by temperament.

But a framework reaching across six domains cannot be verified by someone who is a specialist in none of them, and the audit performed here is only as good as one pair of eyes. So I am asking you to take the chapter nearest your own expertise and put the weight of your field on it. If it gives way, you will have shown me something I could not see, and the Register will then show us both what falls with it. If it holds, it will hold harder for having been pushed, and the framework will be stronger in your hands than it ever was in mine.

That is the collaboration this book needs; it is worth more than a correction to a manuscript. A rational bedrock for the resolution of conflict, if one can be laid at all, will be built by the people who know each domain well enough to test it, and who are willing to carry the parts that survive further than I or any machine could carry them.

I have set out a route to peace as plainly and as thoroughly as I could. I now need a companion traveller — wiser and more knowledgeable than I am — to show me a better route, or to help me strengthen this one.

Premise Register

This Register lists every result the framework establishes, in derivational order, and its rows are grouped under the claim each run supports, under the parts of the General Introduction's arch (§3), so that the Register is the Introduction's table expanded. Each row gives the chapter and section where the result is made, the grade it carries on the four-term ladder, and the results it rests on, cited by chapter and section; where one section holds more than one row, a bracketed tag names the result meant. Principal results — the main conclusions each chapter is built to reach, and the ones later chapters presuppose — are given in bold; the others are the supporting refinements, corollaries, and applications that rest on them.

How to read a dependency. Reject a result in the final column and every row that cites it is exposed. The exposure is result-level, not chapter-level: within an exposed chapter, only the rows resting on the rejected premise fall, and the rest stand untouched. Conditions on a grade are lettered and given beneath the table. A letter is marked on the row whose own claim carries the condition; a row resting on a marked row inherits it without repeating the letter, as Compounding conditions sets out.

The inheritance rule. Self-vindication is a property of a mechanism, and a result is graded self-vindicating only where the mechanism carries it unaided. Where a self-vindicating mechanism runs through a premise graded lower, the result takes the premise's grade. Three rows stand at the top tier: the occurrence premise and the composition premise of Chapter 1 (§5.3), and the foundational result built from the two — that at least one difference obtains. The same result's second route, through the two-names result of Chapter 3 (§3.5), is established at that result's tier; the vindication law and the exhibitive closure of Chapter 7 both run through Chapter 2's thesis and are accordingly established.

Closing and directional results. The grade records how far a result can be leaned on and not what the result does, and two jobs are mixed in the table. A closing result shuts a door: the two-names result (3 §3.5), the exhibitive closure (7 §4.1), the vacancy result (14 §3.6). A

directional result tells the party it meets where to stand instead, at a named cost: the containment result (14 §3.7), which leaves the realist truth-aptness, truth, observer-independence, and corrigibility and removes only the vacant clause; the containment fork (10 §3.2), which leaves the theist the structured ground or the Palamite split; the sustainer (24 §4.1), left standing with his cost itemised; the eminence horn of Chapter 12 (§4.3). The directional results do the work of founding claims by another route; they are the constructive positions a merely diagnostic reading of the framework misses.

Dependency audit and the tracking floor. A row is meant to record what a result uses and not what its chapter assumes, and the rows of the applied half that cite Chapter 14 are audited here against their derivations on the weaker premise the floor names. Three use the floor and no more, and cite it: the unified causal model (18 §2), which needs the responses as forces and not as value; the structure of purpose (19 §2), whose sense is the response whichever way value is read; and the three modes (21 §3), which use transformability and the time-indexed test, the moral character of the preferences (16 §2) fixing the chapter's domain and no premise of the partition. Four survive the refusal of constitution by a substitution the text names: the anchor's observer-independence, which a realist about welfare also holds, carries 16 §6.3, 16 §6.7 and 22 §4; the fit criterion of 19 §6 gives way to the objectivism it approximates (Ch. 19, §6.3); the answer to the relativist at 22 §7 reads the affected party's welfare for her registration (Ch. 22, §7.2); the manipulation line of 21 §7.6 reads accuracy as tracking. One survives by another route: the desert result (20 §6), by Chapter 18 (note q). What needs constitution is what was built to: Chapter 14's own results, the residue of 15 §4.4 (Ch. 15's convergence bet), and the constitutive component of 16 §2 and 16 §4 (note o); the value taxonomy (15 §3.2, 15 §5) keeps its row and survives read as an account of evaluation. The audit covers the rows named; the others record dependencies as their chapters state them.

Grade	Tier	Meaning
SV	Self-vindicating	The denial instantiates what it denies.
E	Established	Validly derived from premises the book defends, so that a challenger must reject a defended premise. The theological entries are established as negative demonstrations, inheriting their security from the constraints they apply.
C	Constructed	The best available positive account: coherent and well-motivated, but not forced, with live rivals remaining.
P	Proposed	An explicitly revisable proposal, offered as the most coherent option that survives the critique.

The Premise Register

Ch. §	Result	Grade	Rests on
Bedrock — Chapter 1: At least one difference obtains: the one thing that cannot be denied.			
1 §5.3 (O0)	**The occurrence premise: when something happens, something really happens**	SV	—
1 §5.3 (O-C)	**The composition premise: an entertaining of the denial is composed of at least two components**	SV	—
1 §3	**Self-vindication of difference**	E[a]	2 §2
1 §5.1	Immunity to Stroud's substitution objection	E	1 §3; 2 §2

Ch. §	Result	Grade	Rests on
1 §5.3	**Ontological standing of difference, from the occurrence and composition premises; neutral between realism and idealism about the constituents of diversity — distinction, not matter, is what later rows draw.**	SV	1 §5.3 (O0); 1 §5.3 (O-C)
1 §5.3 (absence)	Ontological standing of difference by the second route, through the two-names result	E	1 §5.3 (O0); 3 §3.5
Footing: ontological — Chapter 2: Content requires contrast, and the diff is the unit of content.			
2 §2	**Structural dependence of content on contrast**	E[b]	—
2 §3	Structural explanation of the given's dependence: no content is self-standing	E	2 §2
2 §5	Ontological standing of difference, from the asymmetry of the difference predicate	E	2 §2; 1 §5.3 (O0)
Footing: ontological — Chapter 3: The diversity/equality partition is exhaustive.			
3 §2.1	**Formal exhaustiveness of the diversity/equality partition**	—[c]	—

Ch. §	Result	Grade	Rests on
3 §2.3	Substantiveness of the partition, via predicative priority and the constitutive instantiation thesis	E	1 §3; 2 §2; 3 §2.1
3 §2.3 (relational)	The relational corollary: every relation-instance that does structural work has non-identical relata, instantiates distinction, and terminates where distinction terminates	E	3 §2.3; 3 §2.1
3 §3.5	**The two-names result: an undifferentiated condition is equality if it obtains and non-existence if it does not; with the three-contrast criterion**	E^{y}	3 §2.1
3 §3	Exhaustiveness tested against six rival traditions	E	1 §3; 3 §2.1; 3 §2.3; 3 §3.5
3 §4.1	The hard-edge corollary: determinate applicability of the partition	E	3 §2.1; 2 §2
Footing: ontological — Chapter 4: Equality is the foreclosing terminus of grounding.			
4 §5	**Equality as the unique foreclosing terminus of grounding**	E^{d}	2 §2; 3 §2.1; 3 §3.5; 4 §2.2; 4 §3
4 §5 (boundary)	Grounding limitarianism and the unilateral boundary	E^{d}	4 §5; 4 §3
4 §4	The disjunctive ecumenical result	E	3 §2.1; 4 §5

Ch. §	Result	Grade	Rests on
4 §2.2	The two senses of determination	E	3 §3.5; 2 §2
4 §3	The limit-ground / relatum-ground distinction	E	3 §2.3; 4 §2.2
4 §3 (transfer)	The transfer principle: a fact's differentiation is the differentiation of the worldly state it concerns, the articulation of the fact-designator contributing nothing	E	4 §2.2; 3 §3.5
4 App.	The containment order; the unilateral boundary as the non-member bound of a descending chain	E	3 §2.1; 3 §3.5
Footing: ontological — Chapter 5: Time terminates at the same boundary, and the two hierarchies converge.			
5 §2	**Dependence of temporal sequence on distinction; dissolution of the temporal regress**	E	2 §2; 3 §2.1; 4 §5
5 §2.2	The eternal-switch model, and a realist resolution of Kant's first antinomy	E	5 §2; 3 §2.1
5 §4	Convergence of the grounding and temporal hierarchies	E^{d}	4 §5; 5 §2
5 §2.5	The obtaining residue: on the termination branch the series' obtaining is brute, with no feature of its own to ask about	E^{d}	4 §4; 4 §5; 5 §4

Ch. §	Result	Grade	Rests on
Footing: ontological — Chapter 6: The foreclosing terminus is unintelligent.			
6 §2	**The zero-complexity constraint**	E	2 §2; 3 §3.5
6 §2.2	**The equalisation principle**	E	6 §2
6 §3	A structural lower bound on panpsychism	E	6 §2
6 §2.3	Unintelligence of the foreclosing terminus	E[d]	6 §2; 4 §5
Footing: epistemological — Chapter 7: Meaning and difference are one closed grid.			
7 §2	**The biconditional of meaning and difference**	E	2 §2; 1 §3; 3 §2.1
7 §3	The three-grid architecture, its first containment stated as a norm	E[e]	7 §2; 3 §2.1
7 §4.1 (exhibitive)	**Exhibitive closure of the grid of diffs**	E[f]	1 §3; 2 §2; 7 §2
7 §4.1 (ontic)	**Ontic closure of the grid of diffs**	E	7 §2
7 §4	**The structural floor beneath scepticism**	E	1 §3; 7 §4.1 (exhibitive)
Pier: core theology — Chapter 8: “Infinite” names two opposite conditions, and the terminus has three names.			
8 §2	**The three names of the terminus**: zero, one, infinity — the classification and its asymmetry	E	3 §3.5; 4 §5

Ch. §	Result	Grade	Rests on
8 §2.2	Contamination of the physical “singularity” by its henological address	P	8 §2
8 §3	A representational explanation of why the terminus cannot be pictured (the illusion)	E	2 §2; 6 §2; 5 §2
8 §4	**The two infinities**: the distinction between infinite diversity and infinite equality	E	3 §2.1; 6 §2
8 §A.1	Disambiguation of the something-from-nothing question, with its remainder owned	E	3 §3.5; 5 §2.5; 8 §2
8 §A.2	A structural interpretation of the cosmological singularity	P	4 §5; 8 §4
Pier: core theology — Chapter 9: Creation is dependence, and the sustainer survives the transition picture.			
9 §2	**Incoherence of creation pictured as a transition from non-being**	E	3 §3.5; 4 §2.2; 5 §2; 6 §2
9 §2.4	Redescription of creation as transformation	E^{g}	9 §2; 6 §2.3; 10 §3
9 §3	A source of differentiation is not simple, on the causal-adequacy principle the tradition holds	E^{w}	1 §5.3; 10 §3.2 (fork)
9 §5	Disambiguation of Leibniz’s question, with its remainder owned	E	3 §3.5; 9 §2; 5 §2.5; 9 §6.4

Ch. §	Result	Grade	Rests on
9 §6.4	The vehicle premise: a possibility is a way reality could be; its refusal costed	E	2 §2; 3 §3.5
Pier: core theology — Chapter 10: Ground and mind come apart: the simple God is incoherent, the determinate God an inhabitant.			
10 §3	**Separation of explanatory ground from intelligent agent**	E	6 §2; 6 §2.3; 8 §4
10 §3.2 (fork)	The containment fork: a contained form is a determinate content, hence a distinction in the container, or nothing form-specific	E[w]	2 §2; 6 §2; 7 §4.1 (exhibitive)
10 §3 (conflation)	**The ground/mind conflation in the prime mover tradition and its contemporary defences**	E	2 §2; 6 §2; 6 §2.3; 8 §4; 7 §4.1 (exhibitive)
10 §6	The ecumenical result: a ground without mind	E[h]	10 §3; 4 §4
Pier: core theology — Chapter 11: The kalām reaches no necessary being.			
11 §3	**The undischarged burden of the kalām's necessity ascription**	E[i]	3 §3.5; 4 §5; 5 §2.5
11 §4	**The burden asymmetry favouring the framework**	E[i]	11 §3; 5 §2.5

Ch. §	Result	Grade	Rests on
Pier: core theology — Chapter 12: Negative theology names a limit of thought and misreads it as a being.			
12 §3	**The closure argument catches every form of the negative-theological claim**	E	7 §4.1 (exhibitive); 6 §2; 8 §4
12 §4	Negative theology identifies the structural limit of thought but misreads it as a resource; a corollary of 10 §6, the apophatic God correctly read being the structured ground it leaves standing	E	12 §3; 6 §2.3; 10 §6
Pier: core metaethics — Chapter 13: The mental has two axes, content and orientation, and value lies on the second.			
13 §2.2	**The categorical transition principle**	C	2 §2; 6 §2
13 §A.3	**The uniqueness principle**	C	2 §2; 6 §2.2
13 §3	Reinterpretation of four rival theories of consciousness (§3; §A.6)	C	13 §2.2
13 §2.3 (derived)	**The partition of mental features into determination or nothing, derived from the determination dilemma; the tier later chapters draw**	E	2 §2; 6 §2

Ch. §	Result	Grade	Rests on
13 §2.3	**The two axes of the mental: orientation as a determination's only further division (coupling is binary); directional difference as the genus of directive response**	C	2 §2; 6 §2; 13 §2.1
13 §2.1	**The availability thesis**	C[u]	2 §2; 6 §2.2
Pier: core metaethics — Chapter 14: Value is constituted by the evaluative modality's response and anchored in the affected party; the moral is the empathic evaluation of chosen behaviour.			
14 §3	**The response-constitutive account of value**	C	13 §2.3 (derived); 13 §2.3
14 §3 (floor)	The tracking floor: empathy registers moral value, accurately or not, and the moral concerns conduct's bearing on those it affects — what a realist about value who refuses constitution also holds	C[z]	13 §2.3 (derived); 14 §2.2
14 §4 (forms)	The two constitutive forms of goodness, aesthetic and empathic	C	14 §3
14 §5 (morality)	Identification of morality as empathic evaluation of chosen behaviour	C	14 §3; 14 §4 (forms)

Ch. §	Result	Grade	Rests on
14 §A.1	**The naturalist absorption**	C	14 §3.5 (footing); 14 §3.6 (vacancy); 13 §2.3 (derived); 13 §2.3
14 §2.2	The non-teleological standard of accuracy	E[j]	2 §2; 13 §2.3 (derived)
14 §6.4 (directed)	Directedness of evaluative responses	C	13 §2.3 (derived); 2 §2
14 §6.4 (anchor)	**Anchoring of moral disvalue in the affected party's own response**	C	14 §3; 14 §5 (morality); 14 §6.4 (directed)
14 §6.5	The informed-registration anchor	C	14 §6.4 (anchor); 14 §6.4 (directed)
14 §6.4 (fitting)	Registration/orientation analysis of fittingness and wrong-kind-of-reasons	C	13 §2.3 (derived); 14 §6.4 (directed)
14 §A.7	The self-application result	C	7 §2; 14 §3
14 §3.5 (reference)	Closure of the reference route	C	14 §3.5 (footing); 14 §3.6 (vacancy)
14 §3.5 (footing)	The partition footing of the trilemma	E[k]	1 §3; 2 §2; 7 §4.1 (exhibitive); 13 §2.3 (derived)
14 §3.6 (requirement)	The anchoring requirement	E	7 §3

Ch. §	Result	Grade	Rests on
14 §3.6 (residue)	The residue test	E	14 §3.6 (requirement); 2 §2
14 §3.6 (vacancy)	**The vacancy result**	E	14 §3.5 (footing); 14 §3.6 (requirement); 14 §3.6 (residue); 7 §4.1 (exhibitive)
14 §3.7	The containment result	C	14 §3.6 (vacancy); 14 §6.4 (anchor)
14 §3.5 (dominance)	The dominance accounting	C	14 §3.5 (footing); 8 §3
14 §A.2	**The parity result**	E[l]	1 §3; 14 §3.5 (footing)
14 §A.3	**The commitment result**	E	14 §3.6 (vacancy); 7 §2
14 §A.4	**The quarantine of the field**	E[m]	14 §3.5 (footing); 14 §3.6 (vacancy); 14 §A.1; 14 §A.2; 14 §A.3
14 §A.5	**The field and the pull: the premise the externalist refuses modelled and costed, burden on the refuser**	E[v]	14 §3.5 (footing); 14 §3.6 (vacancy); 14 §A.2; 13 §2.3 (derived)
14 §A.8	Survival of the five standard metaethical batteries	C[n]	14 §3; 14 §3.6 (vacancy); 14 §6.4 (anchor)

Ch. §	Result	Grade	Rests on
Pier: core metaethics — Chapter 15: Total value composes primary and added value; across evaluators a trade-off has no fact beyond convergence under mutual registration.			
15 §3.2	**The primary/added value distinction and the structure of evaluation and choice**	C	14 §3; 13 §2.3 (derived)
15 §5	Constitution of the evaluative hierarchy	C	15 §3.2
15 §6	Convergence with rival theories, and answers to the standing challenges	C	15 §3.2; 15 §5
15 §4.4	**Weighing as field-enrichment; across evaluators, convergence under mutual registration**	C	15 §3.2; 14 §3.6 (vacancy); 14 §3.5 (footing)
Pier: core metaethics — Chapter 16: The major ethical traditions mistook conformity to a standard for empathic evaluation.			
16 §2	**That morality is constitutively empathic and other-regarding**	C	14 §3; 14 §5 (morality); 14 §6.4 (anchor)
16 §4	**The great moral conflation common to the major ethical traditions**	C°	16 §2; 13 §2.3 (derived); 15 §3.2
16 §3	The two-layer analysis and its diagnostic reach	C	16 §4; 16 §2
16 §6.3	The moral verdict is fixed by the act’s effect on the affected party	C	14 §6.4 (anchor); 16 §2

Ch. §	Result	Grade	Rests on
16 §6.7	Moral progress as registration-correction	C	14 §6.4 (anchor); 16 §3
16 §2.1	**Reference-fixing of the other-regarding characterisation**	E^p	14 §2.2
16 §2 (mechanism)	**The descriptive, non-prescriptive mechanism statement**	C	16 §2; 14 §3.6 (vacancy)
Curve: applied theology — Chapter 17: Perfection does not entail infallibility; dogma has a four-stage anatomy with a refutation condition.			
17 §2	**The two-dimensional classification of epistemic states**	E	1 §3; 2 §2
17 §4	**The distinction between perfection and infallibility**	E^x	1 §3; 2 §2; 8 §4; 10 §3
17 §6	The four-stage anatomy of dogma	C	17 §2; 17 §4; 16 §4
Curve: applied ethics — Chapter 18: Authorship is bounded: freedom is real from the first-person present and determined from the third-person retrospect.			
18 §2	**The unified causal model, holding for any possible substrate**	C	14 §3 (floor); 15 §3.2; 13 §2.3 (derived)
18 §4	**The perspectival asymmetry**	C	13 §2.2; 18 §2; 15 §5
18 §6	**Bounded authorship, and its consequence for moral responsibility**	C	18 §2; 18 §4

Ch. §	Result	Grade	Rests on
Curve: applied ethics — Chapter 19: A purpose is an object with a sense; it requires no endorsement from outside the field.			
19 §2	**The bipartite structure of purpose**	C	13 §2.3 (derived); 14 §3 (floor); 15 §3.2; 18 §6
19 §4	Its two constitutive forms, aesthetic and empathic	C	19 §2; 14 §4 (forms); 16 §4
19 §6	Division of the meaning-of-life impasse: the structural question resolved, the normative question referred to Chapter 16	C	19 §2; 14 §6.5
Curve: applied theology — Chapter 20: Desert fails by two routes, so restoration replaces retribution.			
20 §2	**That purpose requires neither God nor immortality**	C	19 §2; 9 §2; 15 §3.2
20 §6 (desert)	**The failure of divine retribution's desert basis**	C[q]	18 §6; 14 §3.6 (vacancy); 14 §6.4 (anchor); 14 §A.3
20 §6 (restorative)	The case for restorative over retributive justice	C	20 §6 (desert); 16 §4
20 §6 (universal)	Autonomy-preserving universalism, proposed for the companion God alone as an aim pursued across time rather than a guaranteed outcome	P[r]	20 §6 (desert); 15 §4.4

Ch. §	Result	Grade	Rests on
Curve: applied ethics — Chapter 21: Structural peace comes only through deep persuasion, the transformation of the preferences that generated the conflict.			
21 §3	**The three modes of conflict resolution**	C	15 §3.2; 15 §4.4
21 §4	Their hierarchical relationship	C	21 §3
21 §3.3	Deep persuasion as the only source of structural peace	C	21 §3; 15 §4.4
Curve: applied ethics — Chapter 22: The conditions for deep persuasion are analytic and its institutions means; relativism and absolutism share the authority-dependence that disables lawmaking.			
22 §2	**The structural conditions for deep persuasion: expressibility, receptivity, non-exclusion**	E^s	21 §3.3; 21 §3
22 §4	Equal deliberative standing, distinct from graded moral consideration	C^s	22 §2; 22 §5 (mechanism); 14 §6.4 (anchor)
22 §5 (mechanism)	The lawmaking mechanism converting private hierarchies into shared law	C	15 §3.2; 15 §4.4; 21 §3.3
22 §5 (democracy)	Its democratic identification	P	22 §5 (mechanism)
22 §7	**The diagnosis of bipolar authoritarianism**	C^s	22 §4; 17 §4; 16 §4

Ch. §	Result	Grade	Rests on
Curve: applied theology — Chapter 23: Creed-based judgement is the conflation projected onto earthly standing: the divine wall.			
23 §3	**The divine wall of discrimination**	C	16 §2; 16 §4; 22 §4
23 §4	The three structural transformations required for democratic compatibility	C	23 §3; 17 §4; 16 §2; 10 §3
23 §5	What remains for religion after them	C	23 §4
Keystone — Chapter 24: The God that survives persuades and does not command: the companion.			
24 §4.1	**The two coherent survivors of the critique on the diversity side: the sustainer and the companion**	E	4 §4; 6 §2.3; 9 §2.4; 10 §3; 11 §3
24 §5	**The companion God**	P[t]	24 §4.1; 8 §4; 10 §3; 13 §2.2; 14 §3; 23 §4; 13 §A.3; 5 §2
24 §5.3	Considerations in the proposal's favour, costed as considerations of fit	P	24 §5
24 §A.1	Faith as evaluative orientation toward the transcendent	C	14 §3; 15 §3.2; 16 §2
24 §A.6	Convergence with process theology, by an independent route	C	24 §4.1; 8 §4

[a] Inherits the tier of the structural dependence thesis (Ch. 2); the mechanism runs through it.

[b] The content-level sense anchored: distinguishability certified by composition from diffs, assertion practice entering as evidence only.

[c] Not graded on the ladder: true by logical form, carried by excluded middle, so no grade would add anything.

[d] Conditional on the termination of the hierarchy, and carrying with it the deferred question of whether equality is possible at all (Compounding conditions, below): a row marked d is read over both.

[e] Its first containment stated as a norm — the anchoring requirement Chapter 14 uses (§3).

[f] At the tier of Chapter 2's thesis, through which its mechanism runs.

[g] Established against any view retaining an origination component via theism's own premises; against pure-dependence readings via the agent dilemma (§2.4), inheriting Chapters 6 and 10.

[h] The content-level truthmaker formulation (§4.2) carries a grain condition, licensed by Chapter 2 (§2.4); the containment fork (§3.2) holds on either grain.

[i] Dialectical in the sense Chapter 1 (§2.4) distinguishes: neither result disproves the cause. The upgrade from non-existence to impossibility at §4.2 uses the S5 assumption declared there; the minimal conclusion does not, and the contested-premises list below carries the declaration.

[j] The naturalised-content problem (Millikan, 1984; Dretske, 1988) is recorded as open.

[k] Rests on Chapters 1 and 2, the exhibitive closure of Chapter 7, and the derived tier of Chapter 13's two axes — the determination-or-nothing partition, row 13 §2.3 (derived) — with the transferable-or-not division supplied at §3.5 by logical form; the trilemma therefore uses no result from Chapter 13's constructed tier (the orientation cell and its uniqueness, row 13 §2.3).

[l] Its refutation condition is stated and open.

[m] Established in its negative component; the partition exhaustive by §3.5, the survey of its occupants defeasible, two tolled exits counted, the positive identification retaining its constructed tier.

[n] Frege–Geach, moral twin earth, creeping minimalism, the Darwinian dilemma, and shmagency.

[o] Two components of unequal dependency: the misattribution of bindingness holds on the weaker premise that empathy tracks moral value (§§3.1, 4.4); the stronger charge rests on response-constitution (Ch. 14, §3.5).

[p] Holds on either side of the Turiel–Haidt dispute, by two independent routes (§§2.1, 4).

[q] The desert clause is additionally foreclosed at the level of commitment by Chapter 14 (§A.3); the row stands on two independent routes, and the Chapter 14 route alone carries it for a reader who refuses the sourcehood premise.

[r] Conditional on the survival of persons beyond death and on a divine educator with the means §6 specifies.

[s] Conditional on the aim of aligned social peace (Ch. 21, §1): the necessity claimed is necessity for that aim. The three conditions of 22 §2 follow from what deep persuasion is (Ch. 22, §1), so that row records an analytic consequence of the definition and keeps its grade; 22 §4 and 22 §7 rest on constructed rows — the mechanism, the anchor, the conflation — and carry the constructed grade by inheritance. The chapter's substance is the institutional realisation of the conditions, graded at 22 §5.

[t] Its mindedness is a posit costed rather than derived (§5.1), and its benevolence a posit within the proposal rather than a derivation from structure (§5.2).

[u] Used by no later result: the chapters after 13 draw the determination-or-nothing partition and the existence of valence as the orientation of a content, and a reader who rejects the placement of feltness loses this row and nothing downstream (Ch. 13, §6).

[v] The model, the classification, and the unified costs are established; the jobs table uses the affected-party anchor (14 §6.4), which is constructed, and stands at that tier.

[w] The fork's first horn is refused by the Thomist doctrine of virtual containment, on which a form may be contained without being a distinction in its container; the row stands at the cost Chapter 10

(§3.2) states for that refusal; a challenger who takes the doctrine rejects this row by name.

[x] Two arguments of unequal reach share the row: the evidential argument, on 1 §3 and 2 §2, holds against every guarantor; the structural argument, on 8 §4 and 10 §3, holds against the simple guarantor only; a challenger who refuses simplicity leaves the first intact.

[y] Uses the premise that a lack confers no internal articulation (Chapter 3, §§2.1, 3.5); the realist about negative facts who denies it is met at §3.5; the premise is entered among the contested premises below.

[z] Entailed by the response-constitutive account and held by the realist about value who refuses it: a row citing the floor survives that refusal at the floor's tier, which is the tier of what empathy registers and of the accuracy standard (14 §2.2).

Compounding conditions

The table gives each result one line of dependency, but the conditions of Chapters 1–6 compound, and no single row shows the stack. Three tiers carry them, and a reader auditing any later result should conjoin the tiers it draws on.

The unconditional tier holds the results that hang on no branch: the vindication law and the actual obtaining of at least one difference (Chapter 1), with the modal residue openly conceded — an unheld denial true at a mindless state of pure equality; the dependence of content on contrast, holding on either grain of content-individuation (Chapter 2, §2.4); and the exhaustiveness and determinate applicability of the partition, the latter secured by the hard-edge corollary (Chapter 3, §4.1).

The branch-relative tier is keyed to a single disjunction. The disjunction itself, termination or non-termination, is unconditional; everything on the termination branch is relative to it, and the obtaining residue (Chapter 5, §2.5) confines that branch's entire contingency to a single structureless fact. Rows carrying note *d* sit here.

The deferred tier holds one open question: the metaphysical possibility of equality, shown consistent by Chapter 3's appendix and assigned, with its mirror-image impossibility claim, to the burden

analysis of Chapter 11 (§4.3). Until that adjudication the possibility line of Chapter 4's ledger stands open in both directions.

Appendix A collects the rows that answer the something-from-nothing question — 3 §3.5, 4 §2.2, 5 §2.5, 8 §A.1, 9 §5, and 11 §4 — into one sequence, and adds no row of its own.

Chapter 6's results are compounds and read accordingly: the zero-complexity constraint is derived from the unconditional tier together with the phenomenal-idleness argument (§2.2) and stands with it; the unintelligent terminus adds the termination branch. A result resting on the terminus carries the branch condition and the deferred question together, and the grades are to be read over those conjunctions, never over single links.

Empirical stakes

Nine results are staked, in whole or in part, on tests the chapters specify and have not executed. This list consolidates them so that the exposure can be read in one place, in the spirit in which the Register consolidates the premises; the grades the chapters assign already allow for the fact that none has been run.

Chapter 13 (§A.7). The interface account is staked on phosphene induction under attentional load, probed by no-report measures: a signature surviving withdrawn attention favours the interface account over the workspace account, and its elimination favours the workspace account.

Chapter 14 (§A.6). The placement of the cold verdict as an anchored representation is staked on a divergence between the developmental and the acquired loss of registration, compared at matched vocabulary competence on the moral/conventional distinction on the sensitivity of verdicts to the affected party, and on the force with which verdicts are held: the acquired case ahead on all three is predicted, while the developmental case reaching the acquired case's competence with no history of registration would break the placement.

Chapter 15 (§4.1). The forced-choice diagnostic is staked on convergence under insulated repetition, with the choice-feedback confound controlled: persistent, undiminished reversal under the diagnostic's own conditions would show that nothing is being revealed.

Chapter 15 (§4.4). The convergence of mutually informed fields — each party accurately registering the other's registration — is staked on the contact and deep-canvassing evidence Chapter 21 (§6.2) draws on, with convergence measured as the canvassing designs measure it: a shift of each party's registration toward the other's, after mutual exposure, that persists at delay and survives a counter-message. Fields that converged neither on the outcome nor on the procedure under mutual accurate registration, across a domain, would leave the account's residue wider than it claims there. The contact base is read at the size the experimental re-analysis allows (Paluck, Green, & Green, 2019; Ch. 21, §6.2).

Chapter 18 (§8.9). The inaccessibility claim beneath the perspectival asymmetry is staked on trained introspection: an observer whose reports tracked the subconscious constitution of a specific preference in advance of its arrival, verified against independent measures of that constitution, would tell against the asymmetry, while gains in accuracy about one's own states, which the contemplative literature records, are gains about contents and leave the delivery off-stage.

Chapter 20 (§6). The retributivist's satisfaction as a registration that thins under acquaintance with its object is staked on the restorative-justice comparisons (Sherman & Strang, 2007). A standing desire for the offender's suffering that survived face-to-face acquaintance across those comparisons would falsify the prediction and leave the satisfaction as a registration to be weighed in the mutually informed field.

Chapter 21 (§3.1). The deep/shallow classification of live transformations runs on three prospective axes (concealment, reversion, generalisation) whose verdicts are provisional proxies for durability, the criterion that settles only in retrospect; the axes carry no thresholds, so a classification made on them is a judgement the chapter shows how to make; the stake is that the retrospective criterion will confirm the prospective verdicts more often than not.

Chapter 22 (§A.2). The four institutional predictions await comparative institutional data and are offered as hypotheses; §A.3 records how demanding their falsification conditions are in practice.

Chapter 23 (§3.5). The divine-wall thesis is staked on the doctrinal gradient under the stated controls — ethnicity, minority standing, socioeconomic sorting, and boundary maintenance as such, the last met by comparing traditions and congregations of similar closure that differ in creedal weighting — with a first approximation available from existing datasets; a flat result under those controls falsifies the thesis, with reinterpretation unavailable.

Contested premises

Six premises the framework uses are contested in the literature rather than derived within the book, and a challenger who rejects one of them by name takes a route the table's columns cannot show, since each enters as background to many rows rather than as the result of any. This list names them, states where each is used, and records what its rejection costs and which rows fall with it.

Favouring is grip (the field and the pull). Used at Chapter 14 (§§3.5, 6.1, A.2), Chapter 15 (§6.1), Chapter 16 (§§3.1, 6.4), and Chapter 21 (§6.2); modelled at Chapter 14 (§A.5). The premise holds that a reason is a pull, that a pull needs a body — a register of the evaluative kind — and that a fact which pulls no register is a field, in the physicist's sense, and no reason. The externalist about reasons holds that reasons apply whether or not they bear on any agent, and refuses it. What the refusal costs him: the content of "must," whose only shared model is the one refused, and with it the reference, reliability and exit costs of Chapter 14 (§§3.5–3.6), which the model shows to be one. What falls with it: the parity result (14 §A.2), the refutation condition of the conflation (16 §4.4), and the claim that the rationalist's authority binds no agent in particular (16 §6.4). What stands: the vacancy result on its compositional route, the anchor, the conflation's misattribution component, and the jobs table of Chapter 14 (§A.5), which the anchor delivers without the premise. The affect-free apprehension the externalist offers as evidence for a favouring with no body is placed at Chapter 14 (§A.6) as an anchored representation deployed while the register rests, so the refusal cannot be met with that datum. The sixth-job challenge of §A.5, restated at §A.6 in the externalist's own terms, is recorded as open.

Basic desert requires ultimate sourcehood. Used at Chapter 18 (§§7–8) and Chapter 20 (§§5–6). The premise holds that an agent deserves suffering for a choice only if the self from which the choice issued was her own doing; the compatibilist about desert (Fischer & Ravizza, 1998; Dennett, 2003; Sartorio, 2016) denies it and locates desert in reasons-responsiveness, at the cost argued in that literature rather than here. What falls with it: the authorship route of Chapter 20's case against retribution, and the description of practical responsibility as all that desert leaves. What stands: the retribution result by its value route — the vacancy result, the anchor, and the commitment result of Chapter 14, which use no premise about sourcehood — together with bounded authorship as a description, practical responsibility, and the forward-looking results.

S5, for the kalām's necessity claim. Declared at Chapter 11 (§4.2). The upgrade from "no necessary diverse being" to "a necessary diverse being is impossible" requires that what is possible be necessarily possible, which is S5; the minimal conclusion — that the cause cannot be shown necessary — does not. A challenger who rejects S5 removes only the stronger right-hand conclusion of §4.2's table and leaves rows 11 §3 and 11 §4 untouched.

Labels carry no content beyond what registration supplies (the anchoring requirement). Chapter 7 (§3) read from the hearer's side, with the commitment Chapter 14 (§3.6) states rather than derives — that a content is what can be exhibited or composed, and that registration, being differential causal response (Chapter 13, §2.1), cannot be had of what makes no difference — applied to the framework's own terminus before the realist's; used at Chapter 14 (§3.6) and Chapter 15 (§4.4). The self-standing-circle intuitionist (Parfit, 2011; Scanlon, 2014) holds that a normative vocabulary can stand as arithmetic stands while exiting nowhere into registration, and rejects it; the exit is costed rather than closed, and counted at Chapter 14 (§A.4): a concept with no exhibition and no composition, a faculty that would reach an inert property unaffected by it, and the reliability circle unclosed — the epistemological cost the realist enters on his own account (Enoch, 2011), carried without the mathematical companion (Ch. 14, §3.6). What falls with it: the compositional route of the vacancy result, the residue test, and the commitment result (14 §A.3). What stands: the modal and felt routes of the vacancy result, which leave the inert

favouring without content on the determination dilemma of Chapter 2 (§2.4), and every result outside §3.6. A reader who refuses the residue test here refuses it where Chapter 13 (§2.3) runs it on the hard problem's residue as well, and the Register counts the refusal once.

A lack confers no internal articulation. Used at Chapter 3 (§§2.1, 3.5) by the two-names result and the hard-edge corollary, and through them by every row that rests on 3 §3.5. The premise holds that an absence of distinction, whatever its standing as a fact, introduces no articulation into what lacks it, so that equality bears no internal distinction even where its lack is admitted as a negative fact (Barker & Jago, 2012). The realist about absences who holds that a negative fact articulates its bearer gives equality a feature and empties the equality cell. What falls with the premise: the two-names identity, the termination branch, and Appendix A's answer to the something-from-nothing question. What stands: the formal exhaustiveness, which is logic's; the determination dilemma of Chapter 2; and every result delivered on the non-termination branch.

A possibility is a way reality could be (the vehicle premise). Used at Chapter 9 (§§5, 6.4), Chapter 11 (§6), and Appendix A (§§4.2, 8); named at Chapter 9 (§6.4). The premise holds that what is possible needs a condition to carry it — a way reality could be, actual at itself — so that a modal clause names a candidate state or says nothing. The primitivist about modality (Shalkowski, 1994; Lewis, 2001; Melia, 2005) holds that modal truths need no vehicle and refuses it. What the refusal costs him: a modal clause with no stated subject matter, parted from the open possibility of equality in no stated respect (Ch. 9, §6.4). What falls with it: the modal closure of Leibniz's question, that non-existence was never available (App. A, §4), and row 9 §5 in its modal form. What stands: the factual closure — on neither branch was there ever a state of non-existence (App. A, §5) — the two-names result, and every row that rests on the partition.

The trade-off residue (Chapter 15, §4.4). Not a premise but a limit, recorded beside the premises because every applied verdict inherits it. The account has a theory of what trade-offs are made from and none of the right trade-off: where mutually informed fields converge neither on an outcome nor on a procedure, no further fact settles the conflict, and deliberation is enrichment until the field tips. Nothing falls with it, since it is a limit the account states rather than a

premise it uses; what it qualifies is every verdict in Chapters 15–24 on a conflict of registrations, which is a verdict about what the fields settled on and not about a fact behind them.

The informed-registration residue (Chapter 14, §6.5). Not a premise but a limit, recorded beside the trade-off residue because it qualifies the anchor's content where the anchor's location is fixed. Where a counterfactual registration decides a silent harm, "accurate and complete registration of the object" fixes the moral fact only as far as the completeness it invokes is itself determinate — settled in the easy cases and open in the hard ones. Nothing falls with it; what it qualifies is the content, never the site, of every verdict the informed-registration anchor returns.

Panqualityism (Chapter 6, §3.3). Not a premise but an exposure the zero-complexity constraint cannot assess: a Russellian variant on which qualities are contrastive but undergone by no subject stands outside both the phenomenal-idleness argument, which asks what a quality is like for its bearer, and the equalisation evidence, which needs a field to equalise. Nothing falls with it, since the constraint's downstream uses assume a subject; what stands open is whether the variant is coherent, which the chapter does not adjudicate.

Glossary

This glossary defines the framework-specific terms that recur across chapters, alphabetically. Concepts receive definitions; results are stated as one-sentence claims, their grades carried by the Premise Register, to which any grade mentioned here defers. The chapter reference marks where a term is introduced or most fully developed. Terms confined to a single chapter's local machinery are defined there and not repeated in this list.

Absolute moral responsibility. The claim that a moral agent is fully and unconditionally responsible for her choices — that no mitigating factor (upbringing, neurology, involuntary emotional inputs) qualifies the attribution. Bounded authorship makes absolute moral responsibility unavailable to any possible agent. *See also* practical moral responsibility. (Ch 18, §7.)

Absolute self-refutation. One of the existing kinds of self-refutation in the taxonomy. Occurs when a proposition entails its own negation through logical self-application (e.g. "there are no truths"). Mode-invariant, and needing no holder: the inconsistency is internal to the proposition's logical structure. Blocked by restricting quantifier scope. *See also* pragmatic self-refutation; dialectical self-refutation; self-vindication. (Ch 1, §2.3.)

Absorption result. The closing of the naturalist's escape from the trilemma: "badness just is suffering" either restates the response-constitutive account (absorption) or posits an inert further property the trilemma has already costed. *See also* vacancy result; quarantine of the field. (Ch 14, §A.1.)

Adaptive preference. A preference shaped by the constraints of the agent's situation: sustained by the added value those constraints impose or, where conditioning has transformed the primary response, formed by them outright. The fox who decides the grapes are sour after discovering he cannot reach them. The evaluative hierarchy can be distorted by adaptive preferences, which is why the forced-choice diagnostic must control for situational constraints. (Ch 15, §6.6.)

Added value. Any emotional response elicited by secondary objects associated with a primary one. Added value divides into derived value (flowing backward from the source to the object) and consequential value (flowing forward from the object to its effects). Total value is the composite of primary and added value. (Ch 15, §3.)

Agency. Not a level of the consciousness hierarchy but its output axis: acting is what a system does on what it registers, present from Level 0 and conditioned at each level on that level's field; "choice" names the act at the first level where alternatives are in the field (2, where the internally generated first stands beside the perceived), deliberation the act at Level 6. What a system does is never among the things it registers. *See also* consciousness hierarchy; perspectival asymmetry. (Ch 13, §2.5; Ch 18, §4.2.)

Anchoring (occurrent-badness floor). The regress-stopper of the fittingness account: the moral fact about a party's treatment is constituted by her own registration of her occurrent suffering — observer-independent without being mind-independent. *See also* informed-registration anchor. (Ch 14, §6.4; Ch 16, §6.7.)

Anchoring requirement. The containment of the linguistic grid in the epistemological grid stated as a norm: labels transmit content and registration creates it, so a vocabulary circular at its base must exit into a registered contrast or contribute nothing; read from the hearer's side, a label attached to no grasped diff is no label. The account applies it to its own terminus before the realist's. The value-vocabulary's single exit is valence. *See also* vacancy result; grid of diffs. (Ch 14, §3.6; Ch 7, §3.)

Arch of the argument. The load structure of the twenty-four chapters, drawn in Figure 1: the bedrock (Chapter 1); the footing, ontological (2–6) and epistemological (7); a theological side, pier (8–12) and curve (17, 20, 23); a metaethical side, pier (13–16) and curve (18, 19, 21, 22), whose political crown is Chapters 21–22; and the keystone, peace, set in Chapter 24. Each stone is placed by what its rows rest on, so the arch describes load and never reading order; the

General Introduction sets the claims out along it and the General Conclusion grades them along it. *See also* central claim. (General Introduction, §3; General Conclusion, §1; Premise Register.)

Authority-dependence. The structural defect shared by moral relativism and moral absolutism: both treat authority as a necessary condition for moral legislation. Absolutism demands it from above (a superior source); relativism demands it from below (special standing). Democracy rejects the precondition entirely. (Ch 22, §7.1.)

Autonomy-preserving universalism. The constructive terminus of Chapter 20, proposed for the companion God alone: if persons survive death, the response consistent with bounded authorship is the turning of every soul toward reconciliation by persuasion across time — deep persuasion at its limit, never overwriting — an aim such a God keeps and an outcome he cannot guarantee. The omnipotent educator is a reductio of retribution, not a proposal. *See also* companion God; restorative justice. (Ch 20, §6; Ch 24, §§5.2, 6.5.)

Availability thesis. The identification of feltness with availability to thought: a felt state is content of the kind the percept/thought system can take up, whether or not it is taken up. The consuming system is characterised structurally, without mentioning feltness. The chapter's proposal for the boundary question; used by no later result (Ch 13, §6). (Ch 13, §§2.1, 2.3, 5.2.)

Biconditional (meaningful iff a diff). The foundational biconditional of the grid of diffs: something is meaningful if and only if it is a diff. The biconditional generates the three-grid architecture (ontological, epistemological, linguistic) and establishes the closure of the grid. (Ch 7, §2.)

Bilateral boundary. A terminus at which the grounding relation's constitutive presuppositions are satisfied on both sides — the terminus has features, and the grounded facts above it have features intelligible in light of those features. Schaffer's priority monism posits a bilateral boundary: the cosmos as a maximally complex

terminus. *See also* unilateral boundary, where presuppositions are satisfied on one side only. (Ch 4, §5.)

Bipolar authoritarianism. The shared structure of moral relativism and moral absolutism: a single defect — authority-dependence — disabling the second law of social equality from opposite poles, each pole presenting itself as the other's antidote. The cyclical alternation between the poles within a society is the empirical pattern the diagnosis explains and illustrates, not a result it derives. (Ch 22, §7.)

Blind field. The perceptual analogy for equality. Not the darkness behind closed eyes (which is a visual experience) but the region behind your head — which is neither bright nor dark but absent from experience entirely. The correct experiential analogue for a state without internal distinctions. Used to illustrate why the mind cannot form a positive image of infinite equality. (Ch 8, §3.1.)

Bounded authorship. The account of freedom surviving the unified causal model: the agent really authors her choices, but within an evaluative hierarchy she did not create, using responses she did not command. A musician improvising within a given key. Holds for any possible mind. (Ch 18, §§2–6.)

Brute terminus. A terminus at which explanation is merely declined — a further ground could exist but is not provided. The exit stands tolled: one why-question per feature, the stop unfixed by anything. *See also* limitarian terminus; tolled exit. (Ch 4, §§4–5.)

Categorical transition principle. Increases in mental complexity are constituted by the detection of categorically new kinds of distinction — not by quantitative increases within an existing category. Categories are ordered by structural containment: each higher level presupposes the levels beneath it, and each level from 2 to 7 is derived by one test — presupposing the prior categories while falling outside them — as a new dimension, a new composition, or a new order. (Ch 13, §§2.2, 2.5.)

Central claim. A proposition about the world that a chapter exists to establish, one to a chapter; the twenty-four are set out from bedrock to keystone in the General Introduction (§3) and graded in the General Conclusion (§1). Each is carried by supporting results, which the Premise Register groups beneath it.

Chain architecture. The structural identity shared by intentional and unintentional behaviour: a sequence of attractive and repulsive forces producing an outcome the final link did not author. The unified causal model identifies this shared architecture while preserving the difference between the two cases (the presence or absence of the evaluative modality). (Ch 18, §2.)

Clock of peace. A regulative image for the progressive trajectory of deep persuasion at the social scale. Each true instance of deep persuasion turns the hand forward, expanding the territory of shared understanding. The clock is an ideal, not a prediction: a society may never complete the rotation, but the direction of travel matters. (Ch 21, §4.)

Closure argument. The argument that every cognitive act is a diff and none escapes the grid, holding for any possible mind because it is grounded in the nature of meaning, not human architecture. *See also* exhibitive closure / ontic closure; grid of diffs. (Ch 7, §4.)

Commitment result. The dissolution of the error theory: the response-independence clause marks no contrast, so moral discourse carries no commitment to objective prescriptivity — the commitment Mackie requires is impossible rather than false. Fictionalism inherits the dissolution. *See also* vacancy result; direction projection. (Ch 14, §§3.6, A.3.)

Companion God. A God of infinite diversity within the hierarchy of diversity rather than at its foundation: a companion within existence, not its creator, who lures rather than commands. One of the two conceptions the critique leaves standing on the diversity side, the other being the sustainer; a temporal mind, whose mindedness and benevolent orientation are both posits within the

proposal rather than derivations from diversity. *See also* sustainer (the). (Ch 24, §§4.1, 5, 5.1, 5.2.)

Composition premise. That an entertaining of the denial of difference is composed of at least two components, its mode and its content, with the content negated as a third where the denial is a negation grasped as such and the refusal as a fourth where it is asserted. Self-vindicating: any denial of it is a content taken in some mode. With the occurrence premise it places the ontological result at the top tier. *See also* occurrence premise. (Ch 1, §5.3.)

Conditional love. The structural consequence of combining the claim that God's love is unconditional with the doctrine that eternal punishment awaits those who reject God. If hell is on the table, God's love is conditional on allegiance. True unconditional love cannot include the threat of eternal punishment. (Ch 20, §5.2; Ch 23, §3.3.)

Confidence ladder. The four-term scale grading each result. Self-vindicating: the denial instantiates what it denies. Established: validly derived from defended premises. Constructed: the best available positive account, with live rivals. Proposed: an explicitly revisable proposal. A result may additionally be conditional on a branch, and is at most as secure as its weakest inherited premise. (General Introduction, §5; Premise Register.)

Consciousness hierarchy (Levels 0–7, with Level 4+). The hierarchy of mental complexity constituted by categorical transitions: registration (0), sentience (1), simple consciousness (2), embodied feeling (3), self-consciousness (4; somatic integration as parallel 4+), other-consciousness (5), reflective self-consciousness (6), moral consciousness (7). (Ch 13, §2.5.)

Consequential value. A form of added value that flows forward along the causal chain: from the object to its effects. If a shirt protects sunburnt skin, its consequential value increases. Distinct from derived value. (Ch 15, §3.2.)

Constitutive instantiation argument. The argument form that deploys the constitutive instantiation thesis: any attempt to propose a counterexample to the diversity/equality partition must use the operation (distinguishing) that instantiates the partition's subject matter (distinction), thereby instantiating diversity rather than escaping the partition. Used to establish the exhaustiveness result and invoked in the convergence of hierarchies. (Ch 3, §2.3; Ch 5, §4.)

Constitutive instantiation thesis (CIT). The operation (distinguishing) by which any counterexample to the distinction partition would be proposed instantiates the partition's subject matter (distinction); hence the partition cannot be subordinated to a deeper one and is substantive rather than formal. (Ch 3, §2.3.)

Containment order. The partial order on states by the distinctions they carry: S lies below S′ when every distinction of S is carried by S′ and S′ carries more. Equality is the bottom element and infinite diversity the top; the unilateral boundary is the greatest lower bound, outside the chain, of a descending chain in which no distinction survives. Distinct from the grid containment of Chapter 7 (§3), which orders grids rather than states. (Ch 4, Appendix.)

Content-individuation. The mechanism by which the denial of difference defeats itself: the denial is a determinate content, determinate content requires individuation by contrast, and individuation by contrast is a difference. The mechanism distinguishing self-vindication from the existing kinds of self-refutation. (Ch 1, §§3–4.)

Convergence of hierarchies. The structural finding that the grounding and temporal hierarchies, if both terminate, terminate at the same structural point — equality, zero complexity. The convergence is not coincidental but a consequence of a shared structural dependency on distinction: both hierarchies are ordered by relations that constitutively operate on distinctions, and both encounter the same structural limit at undifferentiation. (Ch 5, §4.)

Cosmic blind field. Equality as it presents when a mind tries to picture "nothing": a state that presents as nothing due to its undetectability while retaining ontological presence — it looks like nothing and is not nothing. Unlike the blind field behind your head, which is contingent and abolished by turning round, it is the blind field of registration as such — any registering of it would require the distinctions it lacks. (Ch 8, §A.1.)

Creation ex nihilo. The doctrine that God brought the universe into being from the complete absence of anything. Chapter 9 argues, from theism's own premise of God's eternity, that origination from non-being is incoherent, while total ontological dependence is coherent but is not origination. *See also* transformation (vs. creation). (Ch 9, §§2–3.)

Creed-based evaluation. The evaluation of persons primarily by their doctrinal allegiance rather than by their character. Orthotic assessment applied to persons — the great moral conflation in its most consequential social form. Contrasted with merit-based evaluation: judgement by character and conduct. (Ch 23, §3.)

Deep persuasion. The mode of conflict resolution in which the conflicting preference itself is changed or removed, by introducing emotionally meaningful insights, perspectives, or experiences that cause the agent to re-evaluate. The only mode that achieves structural peace. *See also* enforcement; shallow persuasion. (Ch 21, §3.3.)

Democratic equality. Equal standing in the legislative process, grounded in the being's actual or potential capacity for valenced states and expressed through the two laws of social equality: all equal under the law, all equal above it. *See also* equal deliberative standing; two laws of social equality. (Ch 22, §4.)

Derived value. A form of added value that flows backward along the causal chain: from the source to the object. A painting's derived value increases when it is identified as a lost Vermeer. Distinct from consequential value. (Ch 15, §3.2.)

Determination dilemma. The forced choice securing the framework's foundational constraints: any putative content either excludes some alternative (and is a diff) or excludes nothing (and specifies nothing); any putative property either is one way rather than another or is no way at all. (Ch 2, §2.4; Ch 6, §2.2.)

Dialectical self-refutation. An existing kind of self-refutation, identified by Castagnoli (2010). Occurs when a thesis-proponent is compelled by the norms of dialectical exchange to concede something inconsistent with the thesis. Mode-dependent: requires a proponent, an interlocutor, and a context of exchange. Blocked by refusing engagement. *See also* pragmatic self-refutation; absolute self-refutation; self-vindication. (Ch 1, §2.4.)

Diff. The minimum unit of determinate content: one of two, individuated by contrast. Any determinate content is a diff — a relational entity existing only in contrast with at least one alternative. Content that contrasts with nothing is the absence of content; the alternative a diff is one-of-two with need not obtain. (Ch 2, §2; previewed Ch 1, §4.1.)

Direction projection. The representational error by which a mind constituted by directional responses represents value as response-independent: unable to represent value undirected, it projects the felt pull into the facts. The metaethical instance of Chapter 8's projection mechanism. (Ch 14, §3.5; Ch 8, §3.4.)

Directional difference. A diff carrying an orientation — a difference felt as a pull rather than merely registered. Its raw form is valence; an emotion is valence bound to content. Aesthetic and empathic goodness are its forms, distinguished by object. (Ch 13, §2.3; Ch 14, §1; Ch 16, §4.1.)

Disjunctive result. The central result of the temporal terminus chapter: either the causal-temporal chain extends infinitely backward as infinite diversity (no temporal origin exists) or it reaches a foreclosing terminus at equality (undifferentiated being at the boundary of time). On neither branch was there ever a state of outright non-existence. (Ch 5, §2.5.)

Diversity. Any state of reality that contains at least one internal distinction — at least one boundary, contrast, or non-identity. Does not require richness or multiplicity; a state with a single internal distinction is already diverse. One of two structural kinds identified by the diversity/equality partition; the other is equality. (Ch 3, §2.1.)

Diversity/equality partition. The exhaustive partition of any state of reality into differentiated (diversity) and undifferentiated (equality). No third option exists, because proposing a third option requires a distinction, and a distinction is diversity. (Ch 3, §2.)

Divine simplicity (doctrine of). The classical doctrine that God has no internal distinction. The framework argues that conjoining absolute simplicity with intelligence is structurally incoherent: cognition requires at least one intra-attribute distinction, which zero complexity excludes. *See also* intra-attribute distinction; identity thesis. (Ch 10, §§1, 4.)

Divine wall of discrimination. The structural consequence of creed-based divine judgement projected onto earthly relationships: a boundary between those on the right side of the doctrinal standard and those who are not. Structurally resistant to legal enforcement because it operates at the level of private evaluative orientation. (Ch 23, §3.)

Dominant purpose. Whatever sits at the top of a person's evaluative hierarchy — the orientation that shapes not just what she chooses but who she is. Dominant purpose is the hierarchy's apex viewed from the perspective of action and identity. It is involuntarily constituted: the agent did not choose which values dominate her hierarchy. (Ch 19, §3.2.)

Doxastic defeat. The grade of self-defeat between assertoric defeat and occupancy elimination: a thesis that resists true belief yet can be truly supposed, such as Moore-paradoxical sentences and Sorensen's blindspots; "nobody believes anything" is one. Falsified by any believing of it, left standing by a supposer. One of the four grades the occupancy test separates. (Ch 1, §6.4.)

Dyology. The framework developed in this book, from the Greek *dyo* (two) and *logos* (account). The name carries the structure at two scales: the minimal two — a difference is one of two, this and what it excludes — and the great two, diversity and equality, the pair that exhausts metaphysical possibility. (Ch 1, §4.1; Ch 3, §2.)

Emotional-pull axis. The second dimension of the two-dimensional classification of epistemic states. For any claim, asks which of two competing thoughts — the thought that the claim could be different, and the thought that it cannot — carries the stronger emotional charge. Its metaphysics is supplied by the evaluative modality account (Ch 14, §3); the classification itself is neutral among accounts of what a charge is. Paired with the logical axis. (Ch 17, §2.2.)

Empathic evaluation. The evaluative modality responding to the emotional consequences of chosen conduct for others. What makes an evaluation moral is its empathic character — not its conformity to a standard. *See also* orthotic assessment. (Ch 16, §2.)

Empathic foundation. The first layer of moral life: the accumulated empathic responses to the emotional consequences of different kinds of conduct, from which moral standards are derived. The evaluative bedrock on which the orthotic superstructure is built. *See also* orthotic superstructure. (Ch 16, §3.)

Enforcement. The mode of conflict resolution in which an agent's behaviour is altered without changing her underlying preference. The attacker is disarmed but may still want to harm you. Addresses survival. *See also* shallow persuasion; deep persuasion. (Ch 21, §3.1.)

Epistemic content. The richer condition in which a state participates in inferential relations — it serves as a reason, supports or undermines other states, and takes its place in the space of reasons. Epistemic content presupposes structural difference (inferential relations require distinguishable relata), but structural difference does not presuppose epistemic content. *See also* structural difference. (Ch 2, §2.1.)

Equal deliberative standing. The standing of a participant in the lawmaking mechanism, grounded in the binary capacity to contribute a perspective for uptake. Ungraded, however much moral consideration varies with what is at stake for the being, which empathic intensity tracks. *See also* democratic equality. (Ch 22, §4.)

Equalisation constraint. The requirement that any adequate theory of consciousness must be compatible with the categorical, cross-modal, configuration-dependent co-elimination of experience and contrast. Rules out theories on which experience is intrinsic to particles, persists without contrast, or diminishes gradually. (Ch 13, §A.5.)

Equalisation principle. Progressive removal of distinctions within any experiential modality produces the structural boundary of experience, not a minimal instance of it: silence, visual silence, numbness. Phenomenological confirmation of the zero-complexity constraint. (Ch 6, §§2.2, 3.)

Equality. Any state of reality containing no internal distinction whatsoever; the word is borrowed to name the opposite of diversity. Ontological one: something exists but nothing within it is distinguishable — being without knowability. Not the social equality of the political chapters, which this sense nearly inverts. *See also* diversity; ontological zero / one / two. (Ch 3, §§2.1, 2.5.)

Eternal switch. A device formalising the possible temporal structures of reality: ON (diversity) and OFF (equality), with no broken position. Sequences containing OFF are diversity-side descriptions; equality bears no ordering relations, so cyclic scenarios describe unordered self-bounded phases. (Ch 5, §§2.2, 2.4.)

Ethereal mind argument. The argument that bounded authorship holds for any possible mind, physical or non-physical. Even an immaterial soul would face the same structural constraint: its evaluative responses must come from somewhere, and the origination of those responses is not itself a product of deliberation. The boundedness is structural. (Ch 18, §5.)

Evaluative field. The totality of emotional responses — to primary and secondary objects — available to a subject at a given moment. Populated by the subconscious jump and constituting the raw material on which deliberation operates. (Ch 15, §2; Ch 21, §3.3.)

Evaluative hierarchy (hierarchy of value). An ordered ranking of objects by the intensity and orientation of the evaluative responses they elicit. Unique to each person, shaped by developmental history, involuntarily constituted, and discovered under forced choice rather than declared in advance. (Ch 15, §5.)

Evaluative modality. The affective capacity through which sentient beings constitute evaluative differences — the oriented differences between better and worse, desirable and repulsive, meaningful and meaningless. Emotion constitutes evaluative differences as vision constitutes chromatic differences. (Ch 14, §3.)

Evaluative orientation. The direction and intensity of a subject's evaluative engagement with a domain, person, or idea. A subject evaluatively oriented toward justice will respond emotionally to injustice even without deliberation. The concept of faith in the companion God chapter is defined as evaluative orientation toward the transcendent. (Ch 15, §2; Ch 24, §A.1.)

Evaluative saturation. The progressive exhaustion of evaluative novelty. After sufficient experience, new encounters fail to generate new evaluative responses because the relevant contrasts have already been registered. The mechanism by which immortality may defeat purpose rather than fulfil it. (Ch 20, §3.)

Exhibitive closure / ontic closure. The two forms of the grid's closure. Exhibitive: no meaning can be exhibited that is not a diff — every counterexample individuates its candidate and thereby grids it. Ontic: no ungridded meaning exists at all, resting on the biconditional. Results operating on attempted sayings route through the exhibitive form alone. (Ch 7, §4.1.)

Faith as evaluative orientation. The reinterpretation of faith as sustained engagement of the evaluative modality with a domain of

experience identified as transcendent, rather than assent to doctrinal propositions. Constituted primarily by primary value — the direct evaluative response — rather than by derived value, the authority of a tradition. (Ch 24, §A.1.)

Falsifier. The fact, supplied by the holding of a self-refuting thesis, that falsifies it: the act performed (pragmatic), the holder's existence (operational), the proposition's own logical form (absolute), a norm of exchange (dialectical), or the individuation of the content (self-vindication). One of the two axes of the taxonomy; two falsifiers are distinct when the same holding supplies by each a different fact falsifying a different thesis. (Ch 1, §2.5.)

Field and pull. The model of "must" both sides of the realist dispute accept, with "field" in the physicist's sense, distinct from the evaluative field: a field is present without a body, a pull only with one, and "must pull" never means pulling with nothing to pull. For reasons, the fact is the field, an evaluative register the body, the reason the pull; a reason that pulls no body is a field, favouring nothing. Refusing the model costs the content of the word. (Ch 14, §A.5.)

Field-enrichment. The mechanism resolving deadlocked evaluation: a deliberative deadlock is an evaluative field in equilibrium, exited not by measurement but by the entry of new secondary objects until the composite tips. Legitimate enrichment must bottom out in real registrations. Across evaluators, enrichment is mutual registration (see mutually informed field). The micro-mechanism of deep persuasion and lawmaking. (Ch 15, §4.4.)

First-person present perspective. The perspective from which the agent is currently evaluating and choosing. From this perspective, freedom is structurally undeniable because the chooser and the choice are identical: disagreement with one's current evaluative orientation would require a different orientation, which would then be the current one. Paired with the third-person retrospective perspective in the perspectival asymmetry. (Ch 18, §4.)

Forced-choice diagnostic (FCD). The method by which the evaluative hierarchy is discovered: a forced choice between valued objects discloses their relative rank, disciplined by insulated repetition and by controls for adaptive preference. (Ch 15, §§4.1, 5.2, 6.)

Four-stage anatomy of dogma. The mechanism by which the conflation of perfection with infallibility produces dogma: (1) earned authority, (2) extended permanence, (3) foreclosed revision, (4) severed foundation. Operates identically across scripture, institutional authority, moral codes, and scientific consensus. (Ch 17, §6.)

Free speech. The institution that secures expressibility — the structural condition that a perspective able to transform a hierarchy can enter the evaluative field. Necessity attaches to the condition; the free-speech right is the best available means of protecting it, defended not from a prior commitment to individual autonomy but from the structural priority of deep persuasion. (Ch 22, §2.)

Great moral conflation. The systematic confusion, across the major ethical traditions, of orthotic assessment (conformity to a standard) with empathic evaluation (response to the emotional consequences of conduct for others). Identified in Kantian, contractualist, divine-command, rule-consequentialist, and virtue-ethical traditions. The ethical hub of the framework's architecture. (Ch 16, §§2–4.)

Grid containment. The chain $G_L \subseteq G_E \subseteq G_O$: every label attaches to a grasped diff, every grasped diff answers to a real distinction, so what can be said sits inside what can be registered, which sits inside what is real. Distinct from the containment order, which orders states rather than grids. (Ch 7, §3.)

Grid of diffs. The three-grid architecture generated by the biconditional "meaningful iff a diff": the ontological grid of real distinctions (G_O), the epistemological grid of grasped diffs (G_E), the linguistic grid of labels (G_L), with containment $G_L \subseteq G_E \subseteq G_O$. At zero complexity all three have no members. (Ch 7, §§2–3.)

Grip. The hold a requirement has on an agent: a felt pull, or a sanction, and nothing else. “Favouring is grip” names the premise that a favouring reaching no register is no favouring, modelled as the field and the pull. *See also* field and pull; parity result. (Ch 16, §3.1; Ch 14, §A.5.)

Ground/mind conflation. The persistent error of treating the demand for an explanatory ground and the demand for an intelligent agent as one demand; the conditions for each are incompatible at the terminus. Aristotle, Aquinas, and Leibniz exhibit the conflation at structurally distinct points. (Ch 10, §3.)

Grounding hierarchy. The ordering of states by the relation of metaphysical dependence: one state obtains because of another. If the hierarchy terminates, equality is its foreclosing terminus — a state of zero complexity at which the grounding relation reaches a unilateral boundary. The framework is ecumenical between termination and non-termination. (Ch 4, §2.)

Grounding limitarianism. The framework’s position on the grounding hierarchy: if grounding terminates, equality is its foreclosing terminus, a unilateral boundary structurally foreclosed rather than brutely terminated, defined in the containment order (Ch 4, Appendix). Distinct from foundationalism, which posits ungrounded diverse facts, and from infinitism, which denies any bound; in the grounding graph’s own terms it is a bounded descent with no minimal element. (Ch 4, §5.)

Hard-edge corollary. The result that the equality cell of the partition is vagueness-proof: whether a state is equality cannot be ontically indeterminate, because a candidate unsettledness about that question is itself a way things stand and therefore disqualifying. Determinate applicability of the partition follows, since with two cells an edge policed on one side is policed on both. (Ch 3, §4.1.)

Identity hypothesis. The hypothesis — motivated but not established by the equalisation constraint — that experience may be constitutively identical with the structurally unique configuration at the interface. Supported by the observation that co-elimination of

experience and contrast is structurally distinct from standard phase-transition dependence. (Ch 13, §A.5.)

Identity thesis (divine simplicity). The simplicity tradition's central strategy: God's attributes are identical with each other and the divine essence. The reply: the thesis collapses inter-attribute distinctions but not intra-attribute ones, which have different truth-conditions and objects. *See also* intra-attribute distinction. (Ch 10, §4.)

Illusion of equality. The systematic misidentification of equality (undifferentiated being) as non-existence (the absence of being). Undifferentiated being appears as nothing without being nothing, because neither has features for detection to latch onto. (Ch 5, §2.5; Ch 8, §3.)

Informed-registration anchor. The account's second anchor, extending the occurrent-badness floor to silent harms: the moral fact is fixed by the party's registration — actual where accurate, counterfactual where not — under a threefold discipline: minimal correction, her modality, diagnostic never overriding. *See also* silent harm; anchoring (occurrent-badness floor). (Ch 14, §6.5.)

Inter-attribute distinction. A distinction between different divine attributes (e.g. between knowledge and power, or between goodness and wisdom). Classical divine simplicity collapses inter-attribute distinctions by identifying all attributes with the divine essence. *See also* intra-attribute distinctions. (Ch 10, §3.)

Interface. The structural boundary where terminus activation carrying a unique triggered signature meets the modality's absence. Consciousness-constituting events are located at this interface, which satisfies all three levels of uniqueness (positional, functional, structural). For the maximal mind of Ch 24 (§5.1) the condition is met in time. (Ch 13, §A.4.)

Intra-attribute distinction. A distinction within a single attribute required for it to be that attribute: knowledge requires known/unknown, will requires willed/unwilled. Survives the

collapse of inter-attribute distinctions; the crux against conjoining simplicity with intelligence. (Ch 10, §4.)

Irrational certainty. The epistemic state in which a claim is logically open, its denial one that can be truly held, but the thought of impossibility carries the stronger emotional charge — the alternative judged closed, or held beyond the reach of evidence. The epistemological basis of dogma. Phenomenologically indistinguishable from rational certainty. (Ch 17, §2.3.)

Kind and token distinctions. Two sorts of diff, both excluded by equality: a kind against its contrary (electron/photon, acid/base) and a token against another token (this region/that region, this moment/the next). Fact-grounding descends toward more token detail and fewer kinds of thing, and the unification of physics is a descent of that shape; values and positions do not thin out, so the base of physics is still diversity. Only equality carries neither. (Ch 4, Appendix.)

Lawmaking mechanism. The operation by which private evaluative hierarchies become shared public standards: through deliberative exchange, new secondary objects enter each participant's evaluative field and restructure the composites until a shared preference emerges and is codified as law. The evaluative architecture of Chapters 14–15 in political deployment; graded constructed, with its democratic identification proposed. (Ch 22, §5.)

Limit-ground / relatum-ground. Two senses of "ground". A relatum-ground stands within the grounding relation, possessing features in virtue of which the grounded obtains; a limit-ground is where the relation's presuppositions fail — where grounding ends, not what performs it. Equality is a limit-ground only. (Ch 4, §3.)

Limitarian terminus. A terminus at which explanation is structurally foreclosed — where the terminating state has no features from which a further explanation could be constructed. Equality, having no internal distinction, cannot ground any explanation of why it obtains. *See also* brute terminus. (Ch 4, §5.)

Logical axis. The first dimension of the two-dimensional classification of epistemic states. For any claim P, asks whether P's negation can be truly held (logically open) or cannot (logically closed). Binary: closed or open. Paired with the emotional-pull axis. (Ch 17, §2.1.)

Modality's absence. The condition that obtains when no activation of a given detection type is present — the structural zero of the relevant modality. Not experienced as emptiness but as the complete absence of the modality. The blind field for vision, silence for hearing, numbness for affect. Distinct from minimal experience within the modality. (Ch 13, §A.4.)

Mode-invariance. The first axis of the taxonomy of self-refutation: whether the defeat survives every change in how the thesis is held — spoken, thought silently, merely supposed — or is blocked by some mode of holding. The operational and self-vindicating forms are mode-invariant; the pragmatic and dialectical forms are mode-dependent. Mode-invariance is the top grade of the occupancy test. (Ch 1, §2.)

Moral technology. A codified standard (rule, principle, norm) that compresses prior empathic evaluations into a teachable, administrable form. Moral technologies are derivative instruments for the orthotic assessment of conduct; their moral authority rests on the empathic evaluations they encode, not on their intrinsic properties. (Ch 16, §3.2.)

Mutually informed field. A field in which each party accurately registers the other's registration, the other's pull entering as an empathic secondary object: the cross-evaluator form of enrichment, produced by mutual registration. A verdict on a trade-off is what mutually informed fields settle on — the outcome where they converge on it, the procedure where only on that, and no further fact where on neither. Convergence is staked empirically. (Ch 15, §4.4.)

Myth of the given. The foundationalist assumption that some epistemic items carry content intrinsically and can stand alone. Sellars: the given cannot justify, lacking inferential connections. The structural diagnosis is prior: the given cannot be self-standing as a content,

since its determinacy is contrast with what it is not, so the foundationalist retreats Sellars leaves open are constrained to present their item as one of two. (Ch 2, §3.)

Non-termination branch. The branch of the disjunctive result on which the temporal hierarchy does not terminate: diversity extends infinitely backward, no temporal origin exists, and the question of what preceded diversity does not arise. Paired with the termination branch. (Ch 5, §2.5.)

Obtaining-determination. The determinateness a state has simply in obtaining rather than not: existential, side-less, requiring no partition within being. Equality has this determinateness and only this, and being non-contrastive it cannot serve as a feature or a second pole. Not a first-order feature (Kant, Frege): obtaining marks a difference only between states, never within one, which is the three-contrast criterion. *See also* property-determination; two-names result. (Ch 3, §3.5; Ch 4, §2.2.)

Obtaining residue. On the termination branch, the framework's total unexplained remainder: the single structureless fact that the bounded hierarchy obtains at all. Which series obtains is a question every rival shares; the residue adds no feature of its own to ask about. *See also* brute terminus. (Ch 5, §2.5.)

Occupancy test. The criterion separating grades of self-defeat by one question: is there any way for the thesis to be both held and true? Dialectical defeat leaves silence; assertoric defeat leaves silent thought; doxastic defeat leaves supposition; occupancy elimination leaves none. The denial of difference is occupancy-eliminated. (Ch 1, §6.4.)

Occurrence premise. When something happens, something really happens: an entertaining is a real occurrence. Self-vindicating in its own right, and the premise three of the four closures of the ontological result draw on (Premise O0). An occurrence stands against its own absence; whatever occurs is diverse. Paired with the composition premise, it carries the ontological result without borrowing. (Ch 1, §5.3.)

Ontological diversity. The conclusion that at least one difference obtains, in whatever hosts the entertaining that denies it — self-vindicating on the occurrence and composition premises together, secured against four retreats by four closures, the whole graded by the occupancy test. The foundation for all downstream results. (Ch 1, §5.3.)

Ontological domain. The domain of states that can coherently obtain. Non-existence is outside the ontological domain: its obtaining would contradict its definition (the absence of obtaining). Equality is inside the ontological domain: it can coherently obtain without contradiction. The distinction between non-existence and equality — both of which present as "nothing" — is the structural basis for the illusion of equality. (Ch 4, §2.1.)

Ontological zero / one / two. A structural analysis of existence and knowledge: at zero (non-existence) nothing to know; at one (undifferentiated being) being without knowability; at two (at least one difference) knowledge first possible. The transition from one to two is the structural origin of the epistemic. (Ch 2, §2.1.)

Open speech. The listener's structural contribution to the persuasive process: the disposition to evaluate a claim on its merits rather than on the identity of the speaker. Its opposite is close-mindedness. Both free speech and open speech are required for deep persuasion. (Ch 22, §3.)

Operational self-refutation. An existing kind of self-refutation, named by Mackie (1964): a proposition refuted by any way of putting it forward, because every way requires a holder whose existence it denies ("I do not exist"). Mode-invariant, so it shares the top grade of the occupancy test with self-vindication; its falsifier is the holder's existence, where self-vindication's is the individuation of the content. Blocked only by removing the holder. (Ch 1, §2.2.)

Orthotic assessment. A cognitive operation that compares an object against a standard and registers conformity or non-conformity. Not an evaluative modality: the operation does not constitutively require emotion. Any emotional response to an orthotic verdict is a separate

aesthetic or empathic response. “Orthotic” derives from the Greek orthos (straight, correct). *See also* empathic evaluation. (Ch 16, §2.)

Orthotic purpose (derivative). A descriptive category in which the agent is oriented toward conformity with a standard. Not constitutively purposive: the standard provides the object of purpose; the aesthetic or (self-)empathic response provides the sense. (Ch 19, §4.2.)

Orthotic superstructure. The second layer of moral life: the set of moral standards, norms, and rules that codify prior empathic evaluations and enable consistency, impartiality, and institutional application. Derivative: its moral authority depends entirely on the empathic foundation from which the standards were derived. *See also* empathic foundation. (Ch 16, §3.)

Parity result. The costing of the normativity objection: the capacity to bind the unmoved belongs to no theory, fact, or foundation, so its absence from the response-constitutive account discriminates against nothing. Paired with the vacancy result it forms a pincer. (Ch 14, §A.2.)

Perfection/infallibility distinction. Perfection is an evidence-based, standard-relative, temporal achievement. Infallibility is a universal claim about all possible evidence. No finite body of evidence can establish a claim about all possible evidence. (Ch 17, §4.)

Perspectival asymmetry. The finding that the first-person present and third-person retrospective perspectives deliver divergent but simultaneously accurate reports of the same choice: freedom phenomenologically real from the first, determination visible from the second. A structural consequence of self-consciousness: the felt openness has an epistemic source, shared with Dennett, and a constitutive one — no standpoint outside the resultant — that survives full information. (Ch 18, §§4, 6.3, 8.5.)

Phenomenal-idleness argument. The derivation of the bridge premise that consciousness requires determination, run in the intrinsicalist’s own vocabulary: a quality indiscernible in presence and absence for

its bearer is phenomenally idle, and phenomenally idle phenomenal character is a contradiction in terms. (Ch 6, §2.2.)

Practical moral responsibility. The form of moral responsibility that bounded authorship preserves. The agent is responsible for what she does with the evaluative resources available to her — for how she exercises bounded authorship — without being responsible for the resources themselves. *See also* absolute moral responsibility. (Ch 18, §7.)

Pragmatic self-refutation. One of the existing kinds of self-refutation: the act of asserting provides a counterexample to the content (shouting "I am not shouting"). Mode-dependent; blocked by removing the act. *See also* operational self-refutation; absolute self-refutation; dialectical self-refutation; self-vindication. (Ch 1, §2.1.)

Predicative priority. The structural property by which distinction is the precondition for all other predication. Every other predicate (colour, shape, causation, temporal sequence, spatial extension) presupposes that the state is already distinguishable from something. One of two independent arguments for the substantiveness of the diversity/equality partition; the other is the constitutive instantiation thesis. (Ch 3, §2.3.)

Presupposition floor. The thesis that the zero-complexity constraint marks the level every quantitative theory of consciousness presupposes. A theory whose quantities are defined over state-spaces assigns nothing to the distinction-free case by grammar rather than by finding, which is why no such theory can disagree with the constraint and none can confirm it. (Ch 6, §5.)

Primary object. Any entity, event, or state of affairs to which the evaluative modality responds directly. The emotional response to a primary object constitutes primary value. (Ch 15, §2.)

Primary value. The emotional response elicited by a primary object in isolation — the modality's direct response to the object itself, apart from any association with its source, effects, or context. Primary

names a route and never a tempo: a fast response may be added, a slow one primary. (Ch 15, §§2, 3.1.)

Primary/added value distinction. The structural finding that total value is a composite of primary value (the evaluative modality's direct response to the object's own features) and added value (derived from source or consequential from effects). The distinction explains why identical objects can differ in value, why value changes without the object changing, and why adaptive preferences have their specific profile. (Ch 15, §3.)

Projection. The representational error by which a mind constituted by distinctions represents what lacks them through what has them: infinite equality imaged as infinite diversity (Ch 8), response-constituted value imaged as response-independent (Ch 14). *See also* representational analysis (of the two-infinities conflation); direction projection. (Ch 8, §3; Ch 14, §3.5.)

Property-determination. The determinateness a state has in bearing a property — a contrastive partition within being (F as against non-F), with both relata available within being. Only diversity has property-determination; equality, being equal throughout, has none. It is the species of determination the zero-complexity constraint concerns: cognition, phenomenal character, and value all require it. *See also* obtaining-determination. (Ch 4, §2.2.)

Pure consciousness event (PCE). A meditative state reported as consciousness without object or content. If it has phenomenal character, that character is a diff within diversity; if none, it is the boundary of consciousness, not an experience. Either way, no counterexample to the zero-complexity constraint. (Ch 6, §4.)

Purpose (evaluative constitution of). Purpose has a bipartite structure: an object (the target) and a sense (the evaluative modality's engagement with it), with two constitutive forms, aesthetic and empathic (including self-empathic); orthotic purpose is derivative. The hierarchy of value generates the dominant purposes that shape identity. (Ch 19, §§2–4.)

Quarantine of the field. The assembly of the trilemma, vacancy, containment, absorption, parity, and commitment results into a closure of the metaethical field. Every stated rival is placed, costed, absorbed, converged, dissolved, contained, or silent, with two tolled exits standing outside the sorting. The premise the externalist refuses is modelled and costed at §A.5. *See also* tolled exit; field and pull. (Ch 14, §A.4.)

Rational certainty. The epistemic state in which a claim is logically closed, its denial one that cannot be truly held, and the thought of impossibility carries the stronger emotional charge. Logic and emotional pull align. Phenomenologically indistinguishable from irrational certainty. (Ch 17, §2.3.)

Relational corollary. The extension of the constitutive instantiation thesis to relations: every relation-instance that does structural work has non-identical relata, therefore instantiates distinction, therefore terminates where distinction terminates. Reflexive instances are conceded and shown idle. (Ch 3, §2.3.)

Representational analysis (of the two-infinities conflation). The explanation of the conflation's persistence: a mind constituted by distinctions cannot form a positive image of a state without them, so every visualisation of infinite equality involuntarily projects diversity onto it. The substitution is invisible from within. (Ch 8, §3.)

Residue test. The instrument separating vacancy from necessary falsity: subtracting a constitutive condition yields contradiction where the term keeps an independent anchor ("heat" keeps its physics), vacancy where the subtracted anchor was its only one — the fate of "value" under the response-independence clause. (Ch 14, §3.6.)

Response-constitutive. The specific character of the evaluative modality account: evaluative differences are constituted by the modality's response, not detected by it. Evaluative properties do not exist independently of the capacity that constitutes them, just as chromatic properties do not exist independently of the visual

system. *See also* response-dependent accounts that merely track pre-existing evaluative facts. (Ch 14, §3.)

Response-dependence. The status of properties constituted by a modality's response to independently existing input: colours to wavelengths, evaluative differences to non-evaluative ones. Response-dependent properties are real; dependence on responses makes them no less so. (Ch 14, §2.)

Restorative justice. The response to wrongdoing that addresses the harm, supports the victim, and transforms the conditions that produced the offence, in place of suffering imposed as deserved. The framework's positive alternative to retribution, divine or human, once retribution's desert basis is removed by two independent routes: bounded authorship, and the absence of any register for which the wrongdoer's suffering is good. *See also* deep persuasion. (Ch 20, §6.)

Richness premise. The premise that grounding many perfections requires internal structure adequate to their plurality. It is denied by the Thomist containment doctrine, and derived within the framework by the containment fork: a contained form is either a determinate content, and so a distinction in the container, or nothing form-specific. (Ch 8, §5; Ch 10, §3.2.)

Secondary object. Any entity, event, or state of affairs associated with a primary object that modifies the total evaluative response. Secondary objects generate added value on top of primary value. (Ch 15, §3.)

Self-standing family. The intuitionist's claim that normative concepts form an irreducibly circular, directly grasped family standing to definition as the logical and mathematical families stand (Parfit, Scanlon). The second of the two tolled exits from the response-constitutive account: costed rather than closed, at the cost of a concept with no exhibition and no composition, a faculty that would reach an inert property unaffected by it, and the reliability circle unclosed; the cost is the epistemological challenge the realist enters on his own account, carried without the mathematical companion

(Ch 14, §3.6). *See also* anchoring requirement; vacancy result. (Ch 14, §§3.6, A.4.)

Self-vindication. A fourth category of self-refutation, irreducible to the pragmatic, operational, absolute, and dialectical kinds: the denial of difference instantiates what it denies through content-individuation rather than through any act of asserting, so no entertaining of the denial, in any mode, is a true one. The vindication law names the mechanism, which runs through Chapter 2's thesis and is graded at its tier. *See also* vindication law; pragmatic self-refutation; absolute self-refutation; dialectical self-refutation. (Ch 1, §3; Ch 2, §2.)

Sense of purpose. The emotional orientation toward an object of purpose — the felt pull that makes the target matter rather than merely being represented. One half of purpose's bipartite structure. *See also* purpose (evaluative constitution of). (Ch 19, §2.1.)

Severing test. The individuation criterion for the primary/added partition, time-indexed: a component of an object's pull is added at a time iff it would not survive severing the association then, intrinsic features held fixed. Transformation of added into primary value is a historical claim, never a synchronic alibi. (Ch 15, §3.2.)

Shallow persuasion. The mode of conflict resolution in which a new, stronger preference is layered over the conflicting preference without removing it. Addresses security but not peace. *See also* enforcement; deep persuasion. (Ch 21, §3.2.)

Silent harm. A wrong in which the affected party's occurrent registration is absent or inverted: the undiscovered betrayal, posthumous wrongs, the blissful violation, aggregation. The cases motivating the informed-registration anchor, which returns the first three verdicts and assigns the fourth to the political layer. (Ch 14, §6.5.)

Something-from-nothing problem. Leibniz's question, "why is there something rather than nothing?", disambiguated. "Nothing" names non-existence, which cannot obtain, so something there cannot fail to be; what the mind pictures in its place is equality,

undifferentiated being, which looks like nothing and is not nothing. The question that survives is why being is differentiated rather than undifferentiated, and its answer is owned as the obtaining residue. *See also* illusion of equality; obtaining residue. (Ch 8, §A.1; Ch 9, §5; App. A.)

Sourcehood premise. That basic desert requires ultimate sourcehood: an agent deserves suffering for a choice only if the self from which the choice issued was her own doing. The line between hard incompatibilism and compatibilism about desert; carried in the Register as a contested premise, and used on one of Chapter 20's two routes against retribution, the other route running through the metaethics alone. (Ch 18, §8.1; Ch 20, §6.)

Stroud's substitution objection. The objection that transcendental arguments establish at most that we must believe certain features obtain. Self-vindication is immune: it operates at the level of content-individuation, and every application of the substitution produces a new instance of the thing in question. (Ch 1, §§5.1–5.2.)

Structural boundary. The condition at which a distinction-dependent relation (grounding, temporal succession, cognition) encounters a state without distinctions. Not a location within the hierarchy but the point at which the hierarchy's own structure terminates. The grounding terminus, the temporal terminus, and the blind field are all structural boundaries. (Ch 4, §5; Ch 5, §2; Ch 6, §2.2.)

Structural conditions for deep persuasion. Expressibility (the perspective can enter the evaluative field), receptivity (it is weighed on its content), and non-exclusion (no relevant perspective is barred in advance). Necessary for deep persuasion at scale; free speech, open speech, and democratic equality are the institutions that secure them — the best available means, not entailments. (Ch 22, §§2–4.)

Structural dependence thesis. Any determinate content must be distinguishable from at least one alternative; content that contrasts with nothing is the absence of content. Secured by the determination dilemma, with the convergence of inferentialist, structuralist, and representationalist theories of content entering as corroboration.

Graded established and left unpromoted: its denial instantiates an instance of the thesis and leaves its universal form untouched. (Ch 2, §2.)

Structural difference. The bare fact that a state is one way rather than another — excluding at least one alternative. Two forms: relational (two items obtain and differ, in quality or at the minimum numerically) and determinative (one item obtains and excludes an alternative that does not); equality has neither. Prior to any category of mind or reality: classifying it already presupposes it. *See also* diff. (Ch 2, §2.1.)

Structural floor. The limit beneath which scepticism cannot descend: the existence of difference. Doubt is constitutively dependent on difference, so doubt cannot coherently target the existence of difference. Distinct from Descartes's cogito, which secures a subject rather than a structural feature of reality. (Ch 7, §4.6.)

Structural foreclosure. The condition at which explanation is not declined but structurally impossible, because the distinctions the grounding relation operates on are absent. Applies wherever a distinction-dependent relation reaches a state without distinctions. *See also* brute terminus; limitarian terminus. (Ch 4, §§4–5.)

Structural foreclosure (temporal). On the termination branch, a diverse series bounded by equality obtains — its obtaining established by the existence of diversity — while why it obtains is unexplainable from the equality side, which has no features from which an explanation could be built. The obtaining question is owned as the obtaining residue. (Ch 5, §§2.3, 2.5.)

Structural peace. Peace achieved by the removal or transformation of the conflicting preference itself, rather than by the suppression of its expression. Only deep persuasion achieves structural peace. Requires alignment at some level under the ceiling principle; the complete alignment of every preference is the limit the clock of peace points toward, never a threshold. Distinct from the cessation of violence, which enforcement can achieve, and the suppression of conflict, which shallow persuasion can achieve. (Ch 21, §§1–2.)

Subconscious hierarchy. The parallel evaluative hierarchy operating beneath awareness. Strong emotional patterns stored in the subconscious — associations with memories, fears, desires, traumas — influence thoughts, preferences, and actions in ways the conscious mind may not recognise. The foundation on which the conscious hierarchy is built. (Ch 15, §5.3; Ch 18, §3.3.)

Subconscious jump. The involuntary mechanism by which the evaluative modality generates an emotional response to an object before deliberate evaluation occurs. The response arrives in consciousness as a fait accompli. The subconscious jump populates the evaluative field that deliberation subsequently operates on. (Ch 18, §3.1.)

Supporting result. A result the book adds in establishing a central claim, graded on its own in the Premise Register on the four-term ladder and listed there beneath the claim it carries; the unit the Register's dependencies are stated in.

Sustainer (the). The other of the two conceptions of God the critique leaves standing on the diversity side: a complex, temporal, non-simple God who is the ground on which everything else depends, creation read as dependence without origination. It keeps the tradition's sovereignty and inherits the problem of evil at full strength. Named and not developed. *See also* companion God. (Ch 24, §4.1; Ch 9, §2.3; Ch 10, §5.1.)

Temporal hierarchy. The ordering of states by causal and temporal succession: the relation "earlier than". Structurally distinct from the grounding hierarchy (which is atemporal) but convergent at the same terminus: if both hierarchies terminate, both terminate in equality. The convergence is explained by the shared structural dependency on distinction: both relations encounter the same structural limit at undifferentiation. (Ch 5, §§2–4.)

Termination branch. The branch of the disjunctive result on which the temporal hierarchy terminates in equality. On it, a diverse series bounded by equality obtains, its obtaining unexplainable from the equality side. Paired with the non-termination branch. (Ch 5, §2.5.)

Third-person retrospective perspective. The perspective on a choice already made or observed in another: the chain of causes is visible, the choice produced by responses the agent did not author. Paired with the first-person present perspective. *See also* perspectival asymmetry. (Ch 18, §4.)

Three modes of conflict resolution. The exhaustive taxonomy of conflict resolution: enforcement (capacity-removal), shallow persuasion (preference-override), and deep persuasion (preference-transformation). The three modes are exhaustive because a preference can be addressed at three levels of the evaluative architecture. Only deep persuasion achieves structural peace. (Ch 21, §3.)

Three names of the terminus. Zero, one, infinity — nothing, the One, the boundless: the three words under which description meets the terminus, individuated by salient contrast (being, multiplicity, limit) and standing to it in non-uniform relations. (Ch 8, §2.)

Three transformations. The three revisions required for religious traditions to be compatible with democratic peace: theocratic to humanistic religiosity; absolutist to democratic religiosity; creed-based to merit-based evaluation. Each independently adoptable. (Ch 23, §4.)

Tolled exit. A position surviving a field's sorting without refutation, held at a cost itemised where the exit is counted. The metaethical field has two; the grounding fork counts the brute terminus the same way. An exit re-enters as a rival only by meeting its cost. (Ch 14, §A.4; Ch 4, §§4, 8.)

Total value. The composite of primary value (the direct evaluative response to an object in isolation) and added value (derived + consequential value contributed by secondary objects). Total value is what the agent actually experiences; primary value is an abstraction recoverable by controlled isolation. (Ch 15, §3.)

Tracking floor. The part of the response-constitutive account that a realist about value also holds: empathy registers moral value,

accurately or not, and the moral concerns conduct's bearing on those it affects. Rows of the Register that cite the floor survive the refusal of constitution; those citing the vacancy result or the constitutive identification do not. *See also* response-constitutive; anchoring (occurrent-badness floor). (Ch 14, §3; the Register's floor row.)

Transfer principle. For the purposes of the grounding hierarchy, a fact's differentiation is the differentiation of the worldly state it concerns: the fact that equality obtains counts as undifferentiated because equality is, the articulation of the fact-designator contributing nothing. Holds on worldly conceptions of facts through obtaining-determination and on structured conceptions through the invariance of the propositional mould. (Ch 4, §3.)

Transformation (vs. creation). The restructuring of already-existing being — in Chapter 9's licensed sense, a diversity-side gloss for the totality-difference between the branch's candidates, never a process with equality as patient. The framework contests only origination from non-being, not transformation. (Ch 9, §§2–4.)

Two axes (content and direction). The exhaustive architecture of the mental: every feature of a mental state is a content or the orientation of one (valence). That every feature is a determination or nothing is derived, and is the tier later chapters use; that a determination's only further division is by orientation — yielding the three cells, determination, orientation of one, or nothing — is constructed; the placement of feltness is the availability thesis's work. (Ch 13, §2.3.)

Two-dimensional classification of epistemic states. The classification generated by crossing a logical axis (can the claim's denial be truly held?) with an emotional-pull axis (which competing thought carries the stronger felt charge?). Yields four states: rational certainty, irrational certainty, rational uncertainty, and irrational uncertainty. The most consequential feature is that rational and irrational certainty are phenomenologically indistinguishable from the inside. (Ch 17, §§2–3.)

Two infinities. The structural ambiguity in "infinity": infinite diversity (endless extension of structure) against infinite equality (complete dissolution of it) — both without limits, in opposite ways. The ambiguity generates the something-from-nothing confusion and the theological conjunction of maximal knowledge with maximal simplicity. (Ch 8, §§2–4.)

Two laws of social equality. The structural conditions for democratic equality. The first law: all are equal under the law — no individual or group is exempt from the standards the society has adopted. The second law: all are equal above the law — no individual or group is excluded from shaping those standards. Violate either and democracy is structurally disabled. (Ch 22, §4.4.)

Two-names result. An undifferentiated condition is equality if it obtains and non-existence if it does not: "equality" and "nothingness" are the two names one featureless condition takes according to whether it is the case, so that a non-existence which obtained would be equality. Cited by later chapters both for the identity it asserts and for the distinction it draws, the two parting only in obtaining. Carries the three-contrast criterion (see obtaining-determination) and uses the premise that a lack confers no internal articulation, entered among the Register's contested premises. (Ch 3, §3.5; Ch 4, §2.1.)

Unified causal model. The constructed account on which intentional and unintentional behaviour share the same chain architecture: a sequence of attractive and repulsive forces producing an outcome the final link did not author. The difference between the two cases is in the kind of force: physical forces govern unintentional behaviour; physical forces plus the evaluative modality's oriented responses govern intentional behaviour. (Ch 18, §2.)

Unilateral boundary. The boundary at which the grounding hierarchy terminates, constituted entirely from the diversity side, as (0, 1) is bounded by 0 without containing it. It is the greatest lower bound, in the containment order, of a descending chain of diverse states that is not itself a member of the chain (Ch 4, Appendix). Schaffer's

monism posits a bilateral boundary; limitarianism a unilateral one, with a featureless obtaining residue against a featured one. (Ch 4, §5.)

Uniqueness principle. If content requires contrast, then what constitutes consciousness must be what makes certain physical events structurally unique. Three hierarchically ordered levels of uniqueness: positional (terminus status), functional (dedicated pathway), and structural (triggered signature). Only events satisfying all three levels at an interface with the modality's absence constitute experience. (Ch 13, §A.3.)

Vacancy result. The result that the robust realist's defining clause — value obtaining independently of every response — secures a contrast by no route: no modal contrast, no felt contrast, no compositional grain. The residue test separates the verdict from necessary falsity. Established on the routes audit and the residue test with one stated commitment (see anchoring requirement). Supports the containment result. (Ch 14, §§3.6–3.7.)

Vehicle premise. The reading of possibility on which a possibility is a way reality could be, actual at itself, so that what is possible needs a condition to carry it. The disambiguation of Leibniz's question rests on it; a primitivist about modality can refuse it, at the cost of a modal clause with no stated subject matter, so the premise is named rather than proved, with the dependency carried in the Register. (Ch 9, §6.4.)

Vindication law. The necessary conditional bearing self-vindication: any possible entertaining of the denial that P applies instantiates a state of affairs to which P applies. Holds at every world, vacuously at mindless ones; any actual entertaining is a vindication instance, never a premise. (Ch 1, §3.1.)

Visual silence. The absence of light — not darkness, which is a visual experience, but the condition in which no visual modality operates. A person born blind does not experience darkness. The correct experiential analogue for a state without internal distinctions. (Ch 6, §2.2; Ch 13, §A.4.)

Zero-complexity constraint. No property whose obtaining constitutively requires at least one distinction can characterise a state with none — knowledge, intention, awareness, agency, design all excluded. The primary hub of the architecture, feeding the theological and consciousness sequences. *See also* equalisation principle. (Ch 6, §2.)

Bibliography

Aaronson, S. (2014). *Why I am not an integrated information theorist (or, the unconscious expander)*. Shtetl-Optimized, 21 May 2014. https://scottaaronson.blog/?p=1799

Abou El Fadl, K. (2001). *Speaking in God's name: Islamic law, authority, and women*. Oneworld.

Adams, M. M. (1999). *Horrendous evils and the goodness of God*. Cornell University Press.

Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley.

Anderson, E. (1999). What is the point of equality? *Ethics*, 109(2), 287–337.

An-Na'im, A. A. (2008). *Islam and the secular state*. Harvard University Press.

Annas, J. (2011). *Intelligent virtue*. Oxford University Press.

Anscombe, G. E. M. (1957). *Intention*. Blackwell.

Appiah, K. A. (2006). *Cosmopolitanism: Ethics in a world of strangers*. Norton.

Arendt, H. (1963). *Eichmann in Jerusalem: A report on the banality of evil*. Viking Press.

Arendt, H. (1970). *On violence*. Harcourt, Brace & World.

Aristotle. (1984b). Nicomachean ethics. In J. Barnes (Ed.), *The complete works of Aristotle* (Vol. 2). Princeton University Press.

Audi, R. (2004). *The good in the right: A theory of intuition and intrinsic value*. Princeton University Press.

Axelrod, R. (1984). *The evolution of cooperation*. Basic Books.

Bach-y-Rita, P., & Kercel, S. W. (2003). Sensory substitution and the human–machine interface. *Trends in Cognitive Sciences*, 7(12), 541–546.

Bain, D. (2013). What makes pains unpleasant? *Philosophical Studies*, 166(S1), 69–89.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, 54(7), 462–479.

Bar-Tal, D. (2013). *Intractable conflicts: Socio-psychological foundations and dynamics*. Cambridge University Press.

Batson, C. D. (2011). *Altruism in humans*. Oxford University Press.

Bayne, T. (2018). On the axiomatic foundations of the integrated information theory of consciousness. *Neuroscience of Consciousness*, 2018(1), niy007.

Beauchamp, M. S., Oswalt, D., Sun, P., Foster, B. L., Magnotti, J. F., Niketeghad, S., Pouratian, N., Bosking, W. H., & Yoshor, D. (2020). Dynamic stimulation of visual cortex produces form vision in sighted and blind humans. *Cell*, 181(4), 774–783.e5.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, 275(5304), 1293–1295.

Becker, E. (1973). *The denial of death*. Free Press.

Bekenstein, J. D. (1981). Universal upper bound on the entropy-to-energy ratio for bounded systems. *Physical Review D*, 23(2), 287–298.
Bell, D. A. (2015). *The China model: Political meritocracy and the limits of democracy*. Princeton University Press.
Benatar, D. (2017). *The human predicament: A candid guide to life's biggest questions*. Oxford University Press.
Benhabib, S. (2004). *The rights of others: Aliens, residents, and citizens*. Cambridge University Press.
Berlin, I. (1969). *Four essays on liberty*. Oxford University Press.
Berman, M. N. (2011). Two kinds of retributivism. In R. A. Duff & S. P. Green (Eds.), *Philosophical foundations of criminal law* (pp. 433–457). Oxford University Press.
Bjørgo, T., & Horgan, J. (Eds.). (2008). *Leaving terrorism behind: Individual and collective disengagement*. Routledge.
Blackburn, S. (1993). *Essays in quasi-realism*. Oxford University Press.
Blair, R. J. R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, 57(1), 1–29.
Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18(2), 227–247.
Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, 30(5–6), 481–499.
Bloom, P. (2016). *Against empathy: The case for rational compassion*. Ecco.
Bonhoeffer, D. (1951/1971). *Letters and papers from prison* (E. Bethge, Ed.). SCM Press.
Borowitz, E. (1991). *Renewing the covenant: A theology for the postmodern Jew*. Jewish Publication Society.
Boyd, R. (1988). How to be a moral realist. In G. Sayre-McCord (Ed.), *Essays on moral realism* (pp. 181–228). Cornell University Press.
Brady, M. S. (2013). *Emotional insight: The epistemic role of emotional experience*. Oxford University Press.
Braithwaite, J. (1989). *Crime, shame and reintegration*. Cambridge University Press.
Bratman, M. E. (1987). *Intention, plans, and practical reason*. Harvard University Press.
Bratman, M. E. (2007). *Structures of agency: Essays*. Oxford University Press.
Brennan, J. (2016). *Against democracy*. Princeton University Press.
Brink, D. O. (1989). *Moral realism and the foundations of ethics*. Cambridge University Press.
Broockman, D., & Kalla, J. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, 352(6282), 220–224.
Broome, J. (1991). *Weighing goods: Equality, uncertainty and time*. Blackwell.
Broome, J. (2013). *Rationality through reasoning*. Wiley-Blackwell.

Buchanan, A., & Powell, R. (2018). *The evolution of moral progress: A biocultural theory*. Oxford University Press.

Buckareff, A. A., & Nagasawa, Y. (Eds.). (2016). *Alternative concepts of God: Essays on the metaphysics of the divine*. Oxford University Press.

Buckareff, A. A., & Plug, A. (2005). Escaping hell: Divine motivation and the problem of hell. *Religious Studies*, 41(1), 39–54.

Burton, J. W. (1990). *Conflict: Resolution and provention*. St. Martin's Press.

Camus, A. (1942/1955). *The myth of Sisyphus* (J. O'Brien, Trans.). Vintage.

Carnap, R. (1932/1959). The elimination of metaphysics through logical analysis of language. In *Logical positivism* (A. J. Ayer, Ed., pp. 60–81). Free Press.

Caruso, G. D. (2021). *Rejecting retributivism: Free will, punishment, and criminal justice*. Cambridge University Press.

Cassam, Q. (2019). *Vices of the mind: From the intellectual to the political*. Oxford University Press.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.

Chalmers, D. J. (2007). Phenomenal concepts and the explanatory gap. In T. Alter & S. Walter (Eds.), *Phenomenal concepts and phenomenal knowledge: New essays on consciousness and physicalism* (pp. 167–194). Oxford University Press.

Chalmers, D. J. (2017). The combination problem for panpsychism. In G. Brüntrup & L. Jaskolla (Eds.), *Panpsychism: Contemporary perspectives* (pp. 179–214). Oxford University Press.

Chalmers, D. J. (2018). The meta-problem of consciousness. *Journal of Consciousness Studies*, 25(9–10), 6–61.

Chang, R. (2002). *Making comparisons count*. Routledge.

Chang, R. (2017). Hard choices. *Journal of the American Philosophical Association*, 3(1), 1–21.

Chen, M. K., & Risen, J. L. (2010). How choice affects and reflects preferences: Revisiting the free-choice paradigm. *Journal of Personality and Social Psychology*, 99(4), 573–594.

Christensen, D. (2007). Epistemology of disagreement: The good news. *Philosophical Review*, 116(2), 187–217.

Christiano, T. (2008). *The constitution of equality: Democratic authority and its limits*. Oxford University Press.

Clarke, R. (2003). *Libertarian accounts of free will*. Oxford University Press.

Clarke-Doane, J. (2020). *Morality and mathematics*. Oxford University Press.

Cleckley, H. (1941/1976). *The mask of sanity* (5th ed.). Mosby.

Cohen, J. (1989). Deliberation and democratic legitimacy. In A. Hamlin & P. Pettit (Eds.), *The good polity: Normative analysis of the state* (pp. 17–34). Basil Blackwell.

Coliva, A. (2015). *Extended rationality: A hinge epistemology*. Palgrave Macmillan.
Corns, J. (2014). The inadequacy of unitary characterizations of pain. *Philosophical Studies*, 169(3), 355–378.
Cottingham, J. (2003). *On the meaning of life*. Routledge.
Craig, W. L. (2008). The absurdity of life without God. In *Reasonable faith* (3rd ed., pp. 65–90). Crossway.
Cuneo, T. (2007). *The normative web: An argument for moral realism*. Oxford University Press.
Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. Putnam.
D'Arms, J., & Jacobson, D. (2000). Sentiment and value. *Ethics*, 110(4), 722–748.
Darwall, S. (2006). *The second-person standpoint: Morality, respect, and accountability*. Harvard University Press.
Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy*, 60(23), 685–700.
de Beauvoir, S. (1947/1976). *The ethics of ambiguity* (B. Frechtman, Trans.). Citadel Press.
Deonna, J. A., & Teroni, F. (2012). *The emotions: A philosophical introduction*. Routledge.
De Sousa, R. (1987). *The rationality of emotion*. MIT Press.
de Waal, F. (1996). *Good natured: The origins of right and wrong in humans and other animals*. Harvard University Press.
Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. Oxford University Press.
Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200–227.
Dennett, D. C. (1984). *Elbow room: The varieties of free will worth wanting*. MIT Press.
Dennett, D. C. (2003). *Freedom evolves*. Viking.
Deutsch, M. (1973). *The resolution of conflict: Constructive and destructive processes*. Yale University Press.
Donaldson, S., & Kymlicka, W. (2011). *Zoopolis: A political theory of animal rights*. Oxford University Press.
Donno, D., & Russett, B. (2004). Islam, authoritarianism, and female empowerment: What are the linkages? *World Politics*, 56(4), 582–607.
Döring, S. A. (2007). Seeing what to do: Affective perception and rational motivation. *Dialectica*, 61(3), 363–394.
Dotson, K. (2014). Conceptualizing epistemic oppression. *Social Epistemology*, 28(2), 115–138.
Dreier, J. (2004). Meta-ethics and the problem of creeping minimalism. *Philosophical Perspectives*, 18(1), 23–44.

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. MIT Press.

Dryzek, J. S. (2000). *Deliberative democracy and beyond*. Oxford University Press.

Duff, R. A. (2001). *Punishment, communication, and community*. Oxford University Press.

Dworkin, R. (1977). *Taking rights seriously*. Harvard University Press.

Dworkin, R. (1996). Objectivity and truth: You'd better believe it. *Philosophy & Public Affairs*, 25(2), 87–139.

Dyson, F. J. (1979). Time without end: Physics and biology in an open universe. *Reviews of Modern Physics*, 51(3), 447–460.

Dyson, F. J. (2001, March 13). Is life analog or digital? *Edge*. https://www.edge.org/conversation/freeman_dyson-is-life-analog-or-digital

Eklund, M. (2017). *Choosing normative concepts*. Oxford University Press.

Elga, A. (2007). Reflection and disagreement. *Noûs*, 41(3), 478–502.

Elster, J. (1983). *Sour grapes: Studies in the subversion of rationality*. Cambridge University Press.

Enoch, D. (2006). Agency, shmagency: Why normativity won't come from what is constitutive of action. *Philosophical Review*, 115(2), 169–198.

Enoch, D. (2011). *Taking morality seriously: A defense of robust realism*. Oxford University Press.

Estlund, D. M. (2008). *Democratic authority: A philosophical framework*. Princeton University Press.

Fine, K. (1994). Essence and modality: The second Philosophical Perspectives lecture. In J. E. Tomberlin (Ed.), *Philosophical perspectives 8: Logic and language* (pp. 1–16). Ridgeview Publishing Company.

Fischer, J. M. (1994). Why immortality is not so bad. *International Journal of Philosophical Studies*, 2(2), 257–270.

Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge University Press.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, 3(4), 552–564.

Fish, M. S. (2002). Islam and authoritarianism. *World Politics*, 55(1), 4–37.

Fishkin, J. S. (2009). *When the people speak: Deliberative democracy and public consultation*. Oxford University Press.

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. MIT Press.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford University Press.

Foot, P. (2001). *Natural goodness*. Clarendon Press.

Fox, K. C. R., Zakarauskas, P., Dixon, M., Ellamil, M., Thompson, E., & Christoff, K. (2012). Meditation experience predicts introspective accuracy. *PLoS ONE*, 7(9), e45370.
Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy*, 68(1), 5–20.
Frankfurt, H. G. (1988). *The importance of what we care about: Philosophical essays*. Cambridge University Press.
Frankfurt, H. G. (2004). *The reasons of love*. Princeton University Press.
Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, 23(11–12), 11–39.
Frankl, V. E. (1946/2006). *Man's search for meaning* (I. Lasch, Trans.). Beacon Press.
Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.
Friedman, J. (2019). Inquiry and belief. *Noûs*, 53(2), 296–315.
Frith, C. D., Blakemore, S.-J., & Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society B*, 355(1404), 1771–1788.
Galtung, J. (1969). Violence, peace, and peace research. *Journal of Peace Research*, 6(3), 167–191.
Garfield, J. L. (2015). *Engaging Buddhism: Why it matters to philosophy*. Oxford University Press.
Geach, P. T. (1965). Assertion. *Philosophical Review*, 74(4), 449–465.
Gibbard, A. (1990). *Wise choices, apt feelings: A theory of normative judgment*. Harvard University Press.
Gibbard, A. (2003). *Thinking how to live*. Harvard University Press.
Goff, P. (2017). *Consciousness and fundamental reality*. Oxford University Press.
Goff, P. (2019). *Galileo's error: Foundations for a new science of consciousness*. Pantheon.
Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1), 20–25.
Goodman, C. (2002). Resentment and reality: Buddhism on moral responsibility. *American Philosophical Quarterly*, 39(4), 359–372.
Greene, J. D. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. Penguin.
Griffin, D. R. (2001). *Reenchantment without supernaturalism: A process philosophy of religion*. Cornell University Press.
Gutiérrez, G. (1971/1973). *A theology of liberation* (C. Inda & J. Eagleson, Trans.). Orbis.
Gutmann, A., & Thompson, D. (2004). *Why deliberative democracy?* Princeton University Press.

Habermas, J. (1990a). Discourse ethics: Notes on a program of philosophical justification. In *Moral consciousness and communicative action* (C. Lenhardt & S. W. Nicholsen, Trans., pp. 43–115). MIT Press.

Habermas, J. (1990b). *Moral consciousness and communicative action*. MIT Press.

Habermas, J. (1996). *Between facts and norms*. MIT Press.

Habermas, J. (2006). Religion in the public sphere. *European Journal of Philosophy*, 14(1), 1–25.

Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience*, 18(4), 196–207.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Pantheon.

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, 65(4), 613–628.

Hardin, C. L. (1988). *Color for philosophers: Unweaving the rainbow*. Hackett.

Harman, G. (1973). *Thought*. Princeton University Press.

Harman, G. (1975). Moral relativism defended. *The Philosophical Review*, 84(1), 3–22.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63(4), 309–321.

Hart, D. B. (2019). *That all shall be saved: Heaven, hell, and universal salvation*. Yale University Press.

Hart, H. L. A. (1961). *The concept of law*. Clarendon Press.

Hartshorne, C. (1948). *The divine relativity: A social conception of God*. Yale University Press.

Hasker, W. (1999). *The emergent self*. Cornell University Press.

Hasker, W. (2004). *Providence, evil and the openness of God*. Routledge.

Hauerwas, S. (1981). *A community of character: Toward a constructive Christian social ethic*. University of Notre Dame Press.

Heidegger, M. (1927/1962). *Being and time* (J. Macquarrie & E. Robinson, Trans.). Harper & Row.

Held, V. (2006). *The ethics of care: Personal, political, and global*. Oxford University Press.

Hick, J. (1989). *An interpretation of religion: Human responses to the transcendent*. Yale University Press.

Hirschman, A. O. (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*. Harvard University Press.

Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge University Press.

Holmes, O. W. (1919). Dissent in *Abrams v. United States*, 250 U.S. 616.
Hookway, C. (2008). Questions, epistemology, and inquiries. *Grazer Philosophische Studien*, 77(1), 1–21.
Horgan, J. (2009). *Walking away from terrorism: Accounts of disengagement from radical and extremist movements*. Routledge.
Horgan, T., & Timmons, M. (1991). New wave moral realism meets moral twin earth. *Journal of Philosophical Research*, 16, 447–465.
Huemer, M. (2005). *Ethical intuitionism*. Palgrave Macmillan.
Hume, D. (1748/1999). *An enquiry concerning human understanding* (T. L. Beauchamp, Ed.). Oxford University Press.
Hume, D. (1751/1998). *An enquiry concerning the principles of morals* (T. L. Beauchamp, Ed.). Oxford University Press.
Hunt, L. (2007). *Inventing human rights: A history*. W. W. Norton.
Hursthouse, R. (1999). *On virtue ethics*. Oxford University Press.
Husak, D. (1992). Why punish the deserving? *Noûs*, 26(4), 447–464.
Inglehart, R., & Norris, P. (2003). The true clash of civilizations. *Foreign Policy*, 135, 62–70.
Ivanhoe, P. J. (2002). *Ethics in the Confucian tradition: The thought of Mengzi and Wang Yangming* (2nd ed.). Hackett.
James, W. (1884). What is an emotion? *Mind*, 9(34), 188–205.
Jeannerod, M. (2006). *Motor cognition: What actions tell the self*. Oxford University Press.
Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, 310(5745), 116–119.
Johansson, P., Hall, L., Sikström, S., Tärning, B., & Lind, A. (2006). How something can be said about telling more than we can know: On choice blindness and introspection. *Consciousness and Cognition*, 15(4), 673–692.
Johns, J. H., & Quay, H. C. (1962). The effect of social reward on verbal conditioning in psychopathic and neurotic military offenders. *Journal of Consulting Psychology*, 26(3), 217–220.
Jonas, H. (1987). The concept of God after Auschwitz: A Jewish voice. *The Journal of Religion, 67*(1), 1–13.
Joravsky, D. (1970). *The Lysenko affair*. Harvard University Press.
Joyce, R. (2001). *The myth of morality*. Cambridge University Press.
Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
Kalanithi, P. (2016). *When breath becomes air*. Random House.
Kalla, J. L., & Broockman, D. E. (2020). Reducing exclusionary attitudes through interpersonal conversation: Evidence from three field experiments. *American Political Science Review*, 114(2), 410–425.
Kane, R. (1996). *The significance of free will*. Oxford University Press.

Kant, I. (1785/1998). *Groundwork of the metaphysics of morals* (M. Gregor, Trans.). Cambridge University Press.

Kant, I. (1788/1997). *Critique of practical reason* (M. Gregor, Trans.). Cambridge University Press.

Kant, I. (1797/1996). *The metaphysics of morals* (M. Gregor, Trans.). Cambridge University Press.

Kauppinen, A. (2012). Meaningfulness and time. *Philosophy and Phenomenological Research*, 84(2), 345–377.

Keller, J. A. (2007). *Problems of evil and the power of God*. Ashgate.

Kelly, T. (2010). Peer disagreement and higher-order evidence. In R. Feldman & T. Warfield (Eds.), *Disagreement* (pp. 111–174). Oxford University Press.

Kelman, H. C. (1958). Compliance, identification, and internalization: Three processes of attitude change. *Journal of Conflict Resolution, 2*(1), 51–60.

Kennett, J., & Fine, C. (2008). Internalism and the evidence from psychopaths and 'acquired sociopaths'. In W. Sinnott-Armstrong (Ed.), *Moral psychology: The neuroscience of morality* (Vol. 3, pp. 173–190). MIT Press.

Kim, J. (1998). *Mind in a physical world: An essay on the mind-body problem and mental causation*. MIT Press.

Kitcher, P. (2011). *The ethical project*. Harvard University Press.

Klein, C. (2015). *What the body commands: The imperative theory of pain*. MIT Press.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, 446(7138), 908–911.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, 100(4), 609–639.

Korsgaard, C. M. (1983). Two distinctions in goodness. *Philosophical Review*, 92(2), 169–195.

Korsgaard, C. M. (1996a). *Creating the kingdom of ends*. Cambridge University Press.

Korsgaard, C. M. (1996b). *The sources of normativity*. Cambridge University Press.

Korsgaard, C. M. (2009). *Self-constitution: Agency, identity, and integrity*. Oxford University Press.

Korsgaard, C. M. (2018). *Fellow creatures: Our obligations to the other animals*. Oxford University Press.

Krauss, L. M., & Starkman, G. D. (2000). Life, the universe, and nothing: Life and death in an ever-expanding universe. *The Astrophysical Journal*, 531(1), 22–30.

Kripke, S. A. (2011). On two paradoxes of knowledge. In *Philosophical troubles: Collected papers* (Vol. 1, pp. 27–51). Oxford University Press.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.
Küng, H. (1971). *Infallible? An inquiry* (E. Quinn, Trans.). Doubleday.
Kvanvig, J. L. (1993). *The problem of hell*. Oxford University Press.
Lackey, J. (2008). *Learning from words: Testimony as a source of knowledge*. Oxford University Press.
Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, 7(1), 12–18.
Landau, I. (2017). *Finding meaning in an imperfect world*. Oxford University Press.
Landemore, H. (2013). *Democratic reason: Politics, collective intelligence, and the rule of the many*. Princeton University Press.
Langton, R. (1993). Speech acts and unspeakable acts. *Philosophy & Public Affairs*, 22(4), 293–330.
Lederach, J. P. (2005). *The moral imagination: The art and soul of building peace*. Oxford University Press.
Leftow, B. (2012). *God and necessity*. Oxford University Press.
Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, 64(4), 354–361.
Levy, N. (2011). *Hard luck: How luck undermines free will and moral responsibility*. Oxford University Press.
Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy*, 61(4), 343–377.
Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, *8*(4), 529–539.
Lichtenstein, S., & Slovic, P. (Eds.). (2006). *The construction of preference*. Cambridge University Press.
List, C. (2019). *Why free will is real*. Harvard University Press.
List, J. A. (2003). Does market experience eliminate market anomalies? *Quarterly Journal of Economics*, *118*(1), 41–71.
Loar, B. (1990). Phenomenal states. In J. E. Tomberlin (Ed.), *Philosophical perspectives 4: Action theory and philosophy of mind* (pp. 81–108). Ridgeview Publishing Company.
Locke, J. (1689/1975). *An essay concerning human understanding* (P. H. Nidditch, Ed.). Clarendon Press.
Lowe, E. J. (2008). *Personal agency: The metaphysics of mind and action*. Oxford University Press.
Lutz, A., Slagter, H. A., Dunne, J. D., & Davidson, R. J. (2008). Attention regulation and monitoring in meditation. *Trends in Cognitive Sciences*, 12(4), 163–169.
Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. Penguin.

Maia, T. V., & McClelland, J. L. (2004). A reexamination of the evidence for the somatic marker hypothesis: What participants really know in the Iowa gambling task. *Proceedings of the National Academy of Sciences*, 101(45), 16075–16080.

Maibom, H. L. (2005). Moral unreason: The case of psychopathy. *Mind & Language*, 20(2), 237–257.

Maitra, I., & McGowan, M. K. (Eds.). (2012). *Speech and harm: Controversies over free speech*. Oxford University Press.

Marquis, D. (1989). Why abortion is immoral. *Journal of Philosophy*, 86(4), 183–202.

Marshall, B. J., & Warren, J. R. (1984). Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet*, 323(8390), 1311–1315.

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 50(4), 370–396.

McDowell, J. (1979). Virtue and reason. *The Monist*, 62(3), 331–350.

McDowell, J. (1985). Values and secondary qualities. In T. Honderich (Ed.), *Morality and objectivity: A tribute to J. L. Mackie* (pp. 110–129). Routledge.

McKenna, M. (2012). *Conversation and responsibility*. Oxford University Press.

McMahan, J. (2002). *The ethics of killing: Problems at the margins of life*. Oxford University Press.

Meiklejohn, A. (1948). *Free speech and its relation to self-government*. Harper & Brothers.

Mele, A. R. (1995). *Autonomous agents: From self-control to autonomy*. Oxford University Press.

Mele, A. R. (2006). *Free will and luck*. Oxford University Press.

Mengzi. (2008). *Mengzi: With selections from traditional commentaries* (B. W. Van Norden, Trans.). Hackett.

Mesle, C. R. (1993). *Process theology: A basic introduction*. Chalice Press.

Metz, T. (2013). *Meaning in life: An analytic study*. Oxford University Press.

Metz, T. (2019). *God, soul and the meaning of life*. Cambridge University Press.

Mill, J. S. (1859/1999). *On liberty* (E. Alexander, Ed.). Broadview Press.

Millikan, R. G. (1984). *Language, thought, and other biological categories: New foundations for realism*. MIT Press.

Moltmann, J. (1972/1974). *The crucified God: The cross of Christ as the foundation and criticism of Christian theology* (R. A. Wilson & J. Bowden, Trans.). SCM Press.

Moody-Adams, M. M. (1999). The idea of moral progress. *Metaphilosophy*, 30(3), 168–185.

Moore, G. E. (1903). *Principia ethica*. Cambridge University Press.

Moore, M. S. (1997). *Placing blame: A theory of the criminal law*. Clarendon Press.
Mørch, H. H. (2019). The argument for panpsychism from experience of causation. In W. Seager (Ed.), *The Routledge handbook of panpsychism* (pp. 269–284). Routledge.
Morris, H. (1968). Persons and punishment. *The Monist*, 52(4), 475–501.
Mouffe, C. (2000). *The democratic paradox*. Verso.
Mouffe, C. (2005). *On the political*. Routledge.
Mouffe, C. (2013). *Agonistics: Thinking the world politically*. Verso.
Murdoch, I. (1970). *The sovereignty of good*. Routledge.
Nagel, T. (1971). The absurd. *Journal of Philosophy*, 68(20), 716–727.
Nagel, T. (1986). *The view from nowhere*. Oxford University Press.
Nelkin, D. K. (2011). *Making sense of freedom and responsibility*. Oxford University Press.
Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. Oxford University Press.
Nietzsche, F. (1887/1998). *On the genealogy of morality* (M. Clark & A. J. Swensen, Trans.). Hackett.
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231–259.
Noddings, N. (1984). *Caring: A feminine approach to ethics and moral education*. University of California Press.
Noggle, R. (1996). Manipulative actions: A conceptual and moral analysis. *American Philosophical Quarterly*, 33(1), 43–55.
Norris, P., & Inglehart, R. (2019). *Cultural backlash: Trump, Brexit, and authoritarian populism*. Cambridge University Press.
Nozick, R. (1974). *Anarchy, state, and utopia*. Basic Books.
Nussbaum, M. C. (2000). *Women and human development: The capabilities approach*. Cambridge University Press.
Nussbaum, M. C. (2001). *Upheavals of thought: The intelligence of emotions*. Cambridge University Press.
Nussbaum, M. C. (2004). *Hiding from humanity: Disgust, shame, and the law*. Princeton University Press.
Nussbaum, M. C. (2006). *Frontiers of justice: Disability, nationality, species membership*. Harvard University Press.
O'Connor, T. (2000). *Persons and causes: The metaphysics of free will*. Oxford University Press.
Oddie, G. (2005). *Value, reality, and desire*. Clarendon Press.
Olson, J. (2014). *Moral error theory: History, critique, defence*. Oxford University Press.
Oord, T. J. (2015). *The uncontrolling love of God: An open and relational account of providence*. IVP Academic.
Oreskes, N. (2019). *Why trust science?* Princeton University Press.

Paluck, E. L., Green, S. A., & Green, D. P. (2019). The contact hypothesis re-evaluated. *Behavioural Public Policy*, 3(2), 129–158.

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. Oxford University Press.

Papineau, D. (2002). *Thinking about consciousness*. Oxford University Press.

Parfit, D. (1984). *Reasons and persons*. Clarendon Press.

Parfit, D. (2011). *On what matters* (Vols. 1–2). Oxford University Press.

Peirce, C. S. (1877). The fixation of belief. *Popular Science Monthly*, 12, 1–15.

Penfield, W., & Rasmussen, T. (1950). *The cerebral cortex of man: A clinical study of localization of function*. Macmillan.

Pereboom, D. (2001). *Living without free will*. Cambridge University Press.

Pereboom, D. (2005). Free will, evil, and divine providence. In A. Dole & A. Chignell (Eds.), *God and the ethics of belief: New essays in philosophy of religion* (pp. 77–98). Cambridge University Press.

Pereboom, D. (2014). *Free will, agency, and meaning in life*. Oxford University Press.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90(5), 751–783.

Pettit, P. (1997). *Republicanism: A theory of freedom and government*. Clarendon Press.

Pew Research Center. (2015). *America's changing religious landscape*. Pew Research Center.

Plantinga, A. (1974). *God, freedom, and evil*. Eerdmans.

Plantinga, A. (1980). *Does God have a nature?* Marquette University Press.

Plantinga, A. (1995). Pluralism: A defense of religious exclusivism. In T. D. Senor (Ed.), *The rationality of belief and the plurality of faith* (pp. 191–215). Cornell University Press.

Plantinga, A. (2000). *Warranted Christian belief*. Oxford University Press.

Plott, C. R. (1996). Rational individual behaviour in markets and social choice processes: The discovered preference hypothesis. In K. J. Arrow, E. Colombatto, M. Perlman, & C. Schmidt (Eds.), *The rational foundations of economic behaviour* (pp. 225–250). Macmillan.

Polger, T. W., & Shapiro, L. A. (2016). *The multiple realization book*. Oxford University Press.

Pollock, J. L. (1986). *Contemporary theories of knowledge*. Rowman & Littlefield.

Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. Routledge.

Prinz, J. J. (2007). *The emotional construction of morals*. Oxford University Press.

Prinz, J. J. (2012). *The conscious brain: How attention engenders experience*. Oxford University Press.

Pritchard, D. (2016). *Epistemic angst: Radical skepticism and the groundlessness of our believing*. Princeton University Press.
Pruss, A. R. (2006). *The principle of sufficient reason: A reassessment*. Cambridge University Press.
Putnam, H. (1975). The meaning of 'meaning'. In *Mind, language and reality: Philosophical papers* (Vol. 2, pp. 215–271). Cambridge University Press.
Putnam, R. D., & Campbell, D. E. (2010). *American grace: How religion divides and unites us*. Simon & Schuster.
Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review*, 60(1), 20–43.
Rabinowicz, W., & Rønnow-Rasmussen, T. (2004). The strike of the demon: On fitting pro-attitudes and value. *Ethics*, 114(3), 391–423.
Railton, P. (1986). Moral realism. *Philosophical Review, 95*(2), 163–207.
Ramachandran, V. S., & Hirstein, W. (1998). The perception of phantom limbs: The D. O. Hebb lecture. *Brain*, 121(9), 1603–1630.
Ratcliffe, M. (2015). *Experiences of depression: A study in phenomenology*. Oxford University Press.
Rawls, J. (1993). *Political liberalism*. Columbia University Press.
Rea, M. C. (2018). *The hiddenness of God*. Oxford University Press.
Regan, T. (1983). *The case for animal rights*. University of California Press.
Repetti, R. (Ed.). (2017). *Buddhist perspectives on free will: Agentless agency?* Routledge.
Roberson, D., Davies, I. R. L., & Davidoff, J. (2000). Colour categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, 129(3), 369–398.
Rosenthal, D. M. (2005). *Consciousness and mind*. Clarendon Press.
Roskies, A. (2003). Are ethical judgments intrinsically motivational? Lessons from 'acquired sociopathy'. *Philosophical Psychology*, 16(1), 51–66.
Sachedina, A. (2001). *The Islamic roots of democratic pluralism*. Oxford University Press.
Sartorio, C. (2016). *Causation and free will*. Oxford University Press.
Scanlon, T. M. (1998). *What we owe to each other*. Harvard University Press.
Scanlon, T. M. (2008). *Moral dimensions: Permissibility, meaning, blame*. Belknap Press of Harvard University Press.
Scanlon, T. M. (2014). *Being realistic about reasons*. Oxford University Press.
Schauer, F. (1982). *Free speech: A philosophical enquiry*. Cambridge University Press.
Scheffler, S. (2003). What is egalitarianism? *Philosophy & Public Affairs*, 31(1), 5–39.
Scheffler, S. (2013). *Death and the afterlife*. Oxford University Press.
Schellenberg, J. L. (1993). *Divine hiddenness and human reason*. Cornell University Press.

Schellenberg, J. L. (2015). *The hiddenness argument: Philosophy's new challenge to belief in God*. Oxford University Press.
Schelling, T. C. (1960). *The strategy of conflict*. Harvard University Press.
Schopenhauer, A. (1839/1985). *On the freedom of the will* (K. Kolenda, Trans.). Blackwell.
Schroeder, M. (2007). *Slaves of the passions*. Oxford University Press.
Schwartz, S. H. (2012). An overview of the Schwartz theory of basic values. *Online Readings in Psychology and Culture*, 2(1), 1–21.
Schwitzgebel, E. (2008). The unreliability of naive introspection. *Philosophical Review*, 117(2), 245–273.
Sen, A. (1999). *Development as freedom*. Knopf.
Seth, A. (2021). *Being you: A new science of consciousness*. Faber & Faber.
Setiya, K. (2017). *Midlife: A philosophical guide*. Princeton University Press.
Shafer-Landau, R. (2003). *Moral realism: A defence*. Oxford University Press.
Sherman, L. W., & Strang, H. (2007). *Restorative justice: The evidence*. The Smith Institute.
Sider, T. (2011). *Writing the book of the world*. Oxford University Press.
Siderits, M. (1987). Beyond compatibilism: A Buddhist approach to freedom and determinism. *American Philosophical Quarterly*, 24(2), 149–159.
Siderits, M. (2003). *Personal identity and Buddhist philosophy: Empty persons*. Ashgate.
Singer, P. (1975). *Animal liberation*. New York Review/Random House.
Singer, P. (1981). *The expanding circle: Ethics and sociobiology*. Farrar, Straus and Giroux.
Slote, M. (2010). *Moral sentimentalism*. Oxford University Press.
Smetana, J. G. (2006). Social-cognitive domain theory: Consistencies and variations in children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 119–153). Lawrence Erlbaum.
Smilansky, S. (2000). *Free will and illusion*. Oxford University Press.
Smith, A. (1759/1976). *The theory of moral sentiments* (D. D. Raphael & A. L. Macfie, Eds.). Clarendon Press.
Smith, M. (1994). *The moral problem*. Blackwell.
Smithies, D. (2019). *The epistemic role of consciousness*. Oxford University Press.
Smuts, A. (2013). The good cause account of the meaning of life. *Southern Journal of Philosophy*, 51(4), 536–562.
Solomon, R. C. (1976). *The passions: Emotions and the meaning of life*. Doubleday.
Sosa, E. (2007). *A virtue epistemology: Apt belief and reflective knowledge* (Vol. 1). Clarendon Press.
Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74(11), 1–29.

Stanley, J. (2005). *Knowledge and practical interests*. Clarendon Press.
Stepan, A. (2000). Religion, democracy, and the "twin tolerations". *Journal of Democracy*, 11(4), 37–57.
Stepan, A., & Robertson, G. B. (2003). An "Arab" more than a "Muslim" electoral gap. *Journal of Democracy*, 14(3), 30–44.
Stout, J. (2004). *Democracy and tradition*. Princeton University Press.
Strawson, G. (1994). The impossibility of moral responsibility. *Philosophical Studies*, 75(1–2), 5–24.
Strawson, G. (2006). Realistic monism: Why physicalism entails panpsychism. *Journal of Consciousness Studies*, 13(10–11), 3–31.
Strawson, P. F. (1962). Freedom and resentment. *Proceedings of the British Academy*, *48*, 187–211.
Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies*, 127(1), 109–166.
Streumer, B. (2017). *Unbelievable errors: An error theory about all normative judgements*. Oxford University Press.
Stump, E. (2010). *Wandering in darkness: Narrative and the problem of suffering*. Oxford University Press.
Sturgeon, N. L. (1985). Moral explanations. In D. Copp & D. Zimmerman (Eds.), *Morality, reason and truth* (pp. 49–78). Rowman & Allanheld.
Sunstein, C. R. (2015). *Choosing not to choose: Understanding the value of choice*. Oxford University Press.
Swinburne, R. (1998). *Providence and the problem of evil*. Clarendon Press.
Swinburne, R. (2004). *The existence of God* (2nd ed.). Clarendon Press.
Tadros, V. (2011). *The ends of harm: The moral foundations of criminal law*. Oxford University Press.
Talbott, T. (2014). *The inescapable love of God* (2nd ed.). Cascade Books.
Tappolet, C. (2016). *Emotions, values, and agency*. Oxford University Press.
Taurek, J. M. (1977). Should the numbers count? *Philosophy & Public Affairs*, 6(4), 293–316.
Taylor, C. (2007). *A secular age*. Harvard University Press.
Taylor, G. J., Bagby, R. M., & Parker, J. D. A. (1997). *Disorders of affect regulation: Alexithymia in medical and psychiatric illness*. Cambridge University Press.
Taylor, R. (1970). *Good and evil*. Macmillan.
Thomson, J. J. (1971). A defense of abortion. *Philosophy & Public Affairs, 1*(1), 47–66.
Tillich, P. (1952). *The courage to be*. Yale University Press.
Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5, 42.
Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17(7), 450–461.

Tsuchiya, N., Wilke, M., Frässle, S., & Lamme, V. A. F. (2015). No-report paradigms: Extracting the true neural correlates of consciousness. *Trends in Cognitive Sciences*, 19(12), 757–770.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.

Tye, M. (1995). *Ten problems of consciousness: A representational theory of the phenomenal mind*. MIT Press.

van Inwagen, P. (2006). *The problem of evil*. Oxford University Press.

Van Norden, B. W. (2007). *Virtue ethics and consequentialism in early Chinese philosophy*. Cambridge University Press.

Vargas, M. (2013). *Building better beings: A theory of moral responsibility*. Oxford University Press.

Volf, M. (2011). *A public faith: How followers of Christ should serve the common good*. Brazos Press.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

Waldron, J. (1999). *Law and disagreement*. Clarendon Press.

Waldron, J. (2006). The core of the case against judicial review. *Yale Law Journal*, 115(6), 1346–1406.

Walls, J. L. (1992). *Hell: The logic of damnation*. University of Notre Dame Press.

Ward, K. (2007). *Divine action: Examining God's role in an open and emergent universe*. Templeton Foundation Press.

Watson, G. (1975). Free agency. *Journal of Philosophy*, 72(8), 205–220.

Wedgwood, R. (2007). *The nature of normativity*. Clarendon Press.

Wegner, D. M. (2002). *The illusion of conscious will*. MIT Press.

Weiskrantz, L. (1986). *Blindsight: A case study and implications*. Clarendon Press.

Whitehead, A. N. (1929). *Process and reality*. Macmillan.

Wielenberg, E. J. (2005). *Value and virtue in a godless universe*. Cambridge University Press.

Wiggins, D. (1980). Deliberation and practical reason. In A. O. Rorty (Ed.), *Essays on Aristotle's ethics* (pp. 221–240). University of California Press.

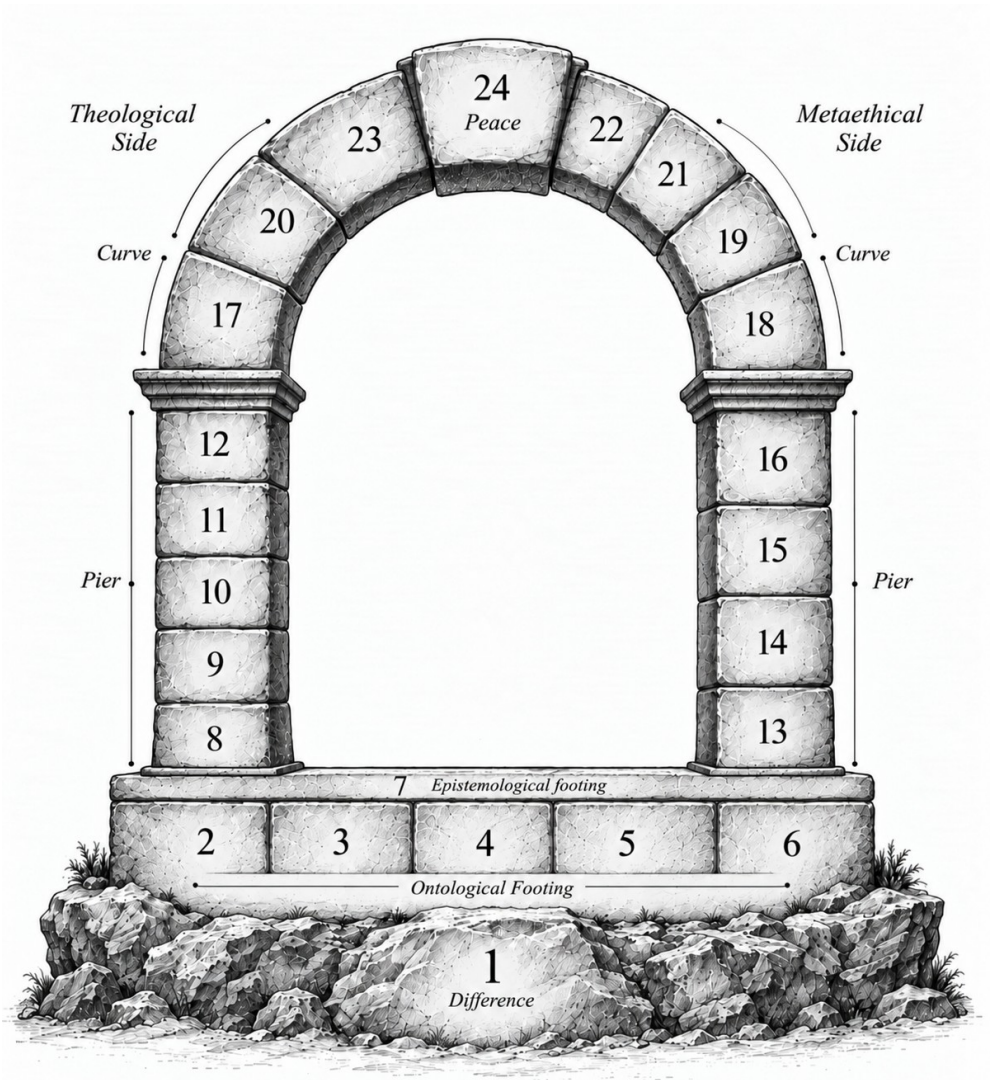
Wiggins, D. (1987). A sensible subjectivism? In *Needs, values, truth: Essays in the philosophy of value* (pp. 185–214). Blackwell.

Williams, B. (1973). The Makropulos case: Reflections on the tedium of immortality. In *Problems of the self* (pp. 82–100). Cambridge University Press.

Williams, B. (1981). Internal and external reasons. In *Moral luck: Philosophical papers 1973–1980* (pp. 101–113). Cambridge University Press.

Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.

Winawer, J., & Parvizi, J. (2016). Linking electrical stimulation of human primary visual cortex, size of affected cortical area, neuronal responses, and subjective experience. *Neuron*, 92(6), 1213–1219.
Wittgenstein, L. (1969). *On certainty* (G. E. M. Anscombe & G. H. von Wright, Eds.; D. Paul & G. E. M. Anscombe, Trans.). Blackwell.
Wolf, S. (2010). *Meaning in life and why it matters*. Princeton University Press.
Wolterstorff, N. (2012). *Understanding liberal democracy: Essays in political philosophy* (T. Cuneo, Ed.). Oxford University Press.
Wolterstorff, N. (2015). *The God we worship: An exploration of liturgical theology*. Eerdmans.
Wong, D. B. (2006). *Natural moralities: A defense of pluralistic relativism*. Oxford University Press.
Xunzi. (2014). *Xunzi: The complete text* (E. L. Hutton, Trans.). Princeton University Press.
Young, I. M. (2000). *Inclusion and democracy*. Oxford University Press.
Zimmerman, M. J. (2001). *The nature of intrinsic value*. Rowman & Littlefield.


Theological Side
Metaethical Side
24
Peace
23
22
21
20
19
Curve
Curve
17
18
12
16
11
15
Pier
10
Pier
9
14
8
13
7
Epistemological footing
2
3
4
5
6
Ontological Footing
1
Difference