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.