



# Undetermined Choices, Luck and the Enhancement Problem

Nadine Elzein<sup>1</sup>

Received: 22 October 2020 / Accepted: 10 October 2021  
© The Author(s) 2021

## Abstract

If indeterminism is to be necessary for moral responsibility, we must show that it doesn't preclude responsibility (the Luck Problem) and that it might enhance it (the Enhancement Problem). A 'strong luck claim' motivates the Luck Problem: if an agent's choice is undetermined, then her mental life will be causally irrelevant to her choice, whichever way she decides. A 'weak luck claim' motivates the Enhancement Problem: if an agent's choice is undetermined, then even if her mental life is causally relevant to her choice, whichever way she decides, we cannot explain how she *settles* her choice. Only the weak luck claim is plausible. However, its plausibility depends on our accepting that we could only settle our choices if they are settled by additional exercises of agency. If we instead understand the process of settling decisions in procedural terms, we can begin to sketch a solution to the Enhancement Problem.

## 1 Introduction

While there is a longstanding worry about reconciling moral responsibility with determinism, it's not obvious that indeterminism fares any better. We could hardly be responsible for something that happens at random.

Some regard this as an irresolvable paradox, rendering free will mysterious (van Inwagen, 2000), or impossible (Waller, 1990). Others are more optimistic about reconciling freedom with determinism than with indeterminism (Hobart, 1934; Smart, 1961; Haji, 2000, 2001; Almeida & Bernstein, 2003).

The challenges for the libertarian are, firstly, to explain how freedom can be rendered *consistent* with indeterminism, and secondly, to explain how (even if we can meet the first challenge) we could justify supposing that indeterminism is *required*

---

✉ Nadine Elzein  
Nadine.Elzein@warwick.ac.uk; nadine@elzein.co.uk

<sup>1</sup> Department of Philosophy, Social Sciences Building, University of Warwick, Coventry CV4 7AL, UK

for freedom. Franklin (2010, 2011) has termed these the ‘Luck Problem’ and the ‘Enhancement Problem’ respectively.

The Luck Problem is not only a challenge for libertarians. It’s a challenge for anyone who supposes that indeterminism would be consistent with freedom (and this includes many compatibilists as well as some ‘hard determinists’). Likewise, the Enhancement Problem is a challenge for anyone who supposes that there is any aspect of freedom that could be more easily reconciled with indeterminism than determinism (and that includes not only libertarians, but some hard determinists, as well as various more complex views, e.g. the view that determinism may threaten some elements of our desert-entailing practices but not others).

I aim to address both problems, not with the goal of vindicating a libertarian perspective, but with that of showing that it’s intelligible to suppose that certain aspects of our desert-entailing practices might fare better if determinism should turn out to be false than they would should it turn out to be true.

In the following section, I will discuss the contrast between the Luck Problem and the Enhancement Problem. I will argue that the Luck Problem only arises if we accept the claim that an indeterministic process is necessarily causally impotent with respect to its outcome, and that we ought to reject this claim. In contrast, the Enhancement Problem rests on the weaker claim that an agent cannot *settle* an outcome if she can exercise no additional control with respect to determining which outcome results from her indeterministic decision-making process. In Sect. 3, I will argue that this objection rests on a particular picture of agential control: one according to which it would have to be grounded in some additional exercise of agency, or else some set of base values that serve as a surrogate for an additional exercise of agency. In the remaining sections, I will sketch an alternative picture of agential control, which appeals to inherent procedural features of the decision-making process itself. This helps us to sketch out a solution to the Enhancement Problem, since it enables us to explain how an agent might have a robust ability to avoid blame, where this counts as a controlled exercise of agency, without our having to posit some additional exercise of agency to ground such control.

## 2 The Luck Problem

There are various arguments for the conclusion that indeterminism precludes free will. Historically, it was argued that an undetermined choice would be ‘uncaused’ or irrational (Collins, 1717; Hobart, 1934; Hume, 2000). Others have argued that the agent ‘disappears’ from any explanation of her own behaviour insofar as events involving the agent do not determine how she chooses (van Inwagen, 1983, 2000; Mele, 1999a; Pereboom, 2001, 2004, 2014; Haji, 2000; Almeida & Bernstein, 2003; See also: Hobart, 1934; Smart, 1961). Others suppose that indeterminism rules out freedom because it entails that there is no ‘contrastive explanation’ for the agent’s choice; no causal explanation of why the agent acted as she did *as opposed to* doing something else (Mele, 1999b, 2005, 2006; Almeida & Bernstein, 2003. See also Haji, 2000, 2001).

I will focus on the Disappearing Agent formulation. The accusation that undetermined choices<sup>1</sup> would be ‘uncaused’ is rare in contemporary writing, as probabilistic accounts of causation are now commonplace. It is similarly rarely argued now that undetermined choices would necessarily be irrational; it’s widely conceded that an agent’s reasons might support rival incompatible courses of action. Arguments that rest on a demand for contrastive explanations are controversial (see Elzein, 2019), and the considerations that favour endorsing such a demand are closely related to the worries that underpin the Disappearing Agent argument in any case. Namely, worries about how agents (or their mental states) could have the right sort of explanatory role in relation to their choices and actions if they do not deterministically fix them.

If the occurrences in an agent’s mental life do not fix the outcome of her deliberation, then we might wonder how she could possibly be responsible for this outcome. Since her own mental states do not explain what she does, it seems to be a matter of chance. How could anyone could be in control of something that happens just by chance?

The problem seems especially (some suppose exclusively) troubling for event causal libertarians. I will address the problem with a broadly event causal picture in mind.<sup>2</sup> Event causalists argue that responsible choices must be causally explained by the right sorts of events involving the agent (events connected to her desires, values, beliefs, etc.) Proponents of the luck objection argue that this requirement cannot be met with respect to a choice which is not causally determined. An undetermined choice would not be explained by occurrences in the agent’s mental life, but by chance.

As Almeida and Bernstein put it, ‘once the indeterminism sets in—be it as a product of motives and character or by the spinning of a roulette wheel—the agent disappears from the scene. After the moment that the indeterminism is produced in the agent, the agent loses any influence that he allegedly had’ (2003, pp. 99–100).

The idea that agents (or events in their mental lives) lose their influence if their choices are undetermined, however, is ambiguous. We can separate a strong reading from a weak one:

- Strong:* If an agent’s choice is undetermined, then events involving the agent could not causally influence her subsequent decision, whichever way she decides.
- Weak:* If an agent’s choice is undetermined, then while events involving the agent may well causally influence her subsequent decision, whichever way she decides, they could not causally influence how she settles which of those decisions to make.

<sup>1</sup> I will, throughout, use the word ‘choice’ to indicate the process of deciding itself, as opposed to the choice that an agent *faces* between options when she makes a decision.

<sup>2</sup> My analysis appeals to processes as opposed to events, but it’s certainly ‘event causal’ as contrasted with the ‘substance causal’ picture.

The Luck Problem (unlike the Enhancement Problem) aims to establish not only that indeterminism doesn't enhance freedom, but that it positively *precludes* it. It seems (a) that it's the strong claim that would need to be supported if determinism is to positively preclude freedom, and (b) that this claim is false.

Let's begin with (a). Suppose only the *weak* luck claim is true—that there is an adequate rational basis for the agent's choice whichever way she decides, and that either way, her choice is caused by events suitably connected to her values, beliefs, desires, etc. Why should the mere possibility of deciding otherwise be freedom-undermining, if the right sort of causal history would be present whichever way she chooses? Kane (1996) notes that if an agent deliberately does something, the mere *possibility* of failing to do it would not plausibly undermine her responsibility.

If there is an adequate basis for her choice whichever way she decides, it's unclear why she should be any less responsible than she would be if her choice were causally determined. Even if her choice *is* determined, there may be rival considerations that count against her settling it as she does. The mere possibility that these might have stopped her from deciding this way seems a dubious basis upon which to exculpate her from responsibility. The weak luck claim alone seems insufficient to establish that indeterminism precludes responsibility. It is the strong luck claim that would need to be defended.

That brings us to (b). The strong luck claim is false. To see why, it helps to think in terms of the causal role of *processes* rather than events. Events are often conceptualised as bounded course-grained units, strung together by chancy or deterministic causal relations, which determine the chances of one discrete event bringing about another. But this picture is misleading.

Suppose we instead think of decision-making as a process. Processes are continuous and fine-grained. If a process is indeterministic, this describes the manner in which it unfolds; it does not describe its weakened causal influence over what follows. When we think of a 'decision' as undetermined, we are apt to imagine that it may or may not occur *even* if we hold the agent's decision-making process fixed. But this is the wrong way to think about it. A decision-making process is *indeterministic* if it could have unfolded differently. In contrast, it's *causally effective* if the way that it unfolds shapes the way that the agent decides.

This reflects a view of causation according to which causal influence depends on difference-making relations between variables (where these measure *features* or *states* of processes or substances). This view stands opposed to an outdated view that construes causation as an all-or-nothing 'bringing about' relation between discrete events. (See Lewis (2000), Paul (2000), Hitchcock (2001), Woodward (2003), Woodward and Hitchcock (2003), Fenton-Glynn (2017)).

We should not think of choices as the chancy upshots of preceding events, but as the culmination of chancy processes. Think again of Almeida and Bernstein's claim, quoted earlier: "After the moment that the indeterminism is produced in the agent, the agent loses any influence that he allegedly had" (2003, pp. 99–100). It is clear that the agent's influence, on this picture, is supposed to be exhausted by occurrences that happen prior to the choice itself, aimed at causally producing it. Indeterminism, on this picture, is not a way of describing the decision-making process itself, but a sort of external disruption, located after the agent's input, that confuses

its output; as if the agent contributes all that she can, and then chance simply ‘takes over’ and fixes the outcome, irrespective of anything *she* has contributed. Hence her own efforts appear causally impotent.

This picture implicitly locates the chanciness *outside* of the decision-making process, creating the impression that it’s an external determinant of the agent’s behaviour; that chance either competes with the process in order to fix the outcome, or entirely supplants it.

Consider an analogy. Suppose an appointing committee is deliberating about which candidate ought to be offered a job, and that the outcome is not causally determined. There are various stories we might tell about this process and its effects. Here’s one story: The committee will deliberate and reach a decision, but the company director is corrupt and nepotistic, so she’ll probably ignore their recommendation and offer the job to her niece. Or perhaps she is not corrupt, merely capricious, and will pick a candidate at random, regardless of the committee’s recommendation.

In this story, the ‘chanciness’ is located *between* the committee’s deliberation and the outcome. This constitutes a weakening of the committee’s causal influence. Even if the director *happens* to offer the job to the candidate that the committee chooses, the committee’s deliberation seems causally impotent.

But the stipulation that the outcome is not causally determined, by itself, need *not* entail that the committee’s deliberation is causally impotent. If the process of decision-making unfolds in an indeterministic manner, this is no barrier to the process causally fixing the outcome. Suppose that the candidates were all very strong and choosing between them was difficult. It may be that the deliberation process could have unfolded in different ways, and that had the process unfolded differently, another candidate would have been chosen.

This differs from the case in which the head of the company essentially acts as a randomiser, fixing the outcome regardless of what the committee advises. Rather, the indeterminism is part of the intrinsic structure of the appointing process itself, and not something that ‘competes’ with the process, or disrupts its power to fix the outcome. While it is common for proponents of the luck problem to draw an analogy with cases in which the agent’s choice is ultimately fixed by a randomising mechanism, disconnected from her deliberation,<sup>3</sup> such analogies are misplaced once we see choices as the culminations of decision-making processes, and not as mere causal outputs of fixed earlier events.

The strong luck claim rests on the assumption that decision-making must be understood as analogous to the case with our capricious company director, when it more closely parallels the case in which the committee’s appointing process is indeterministic.

The view defended here echoes that defended by McCall and Lowe (2005), and Lemos (2018, pp. 106–112), who also utilise a process view of decision-making in

<sup>3</sup> Mele likens the presence of indeterminism within an agent’s psychology to a randomising device inside her brain (Mele, 1999a, p. 277. See also: 2014, p. 554). Pereboom imagines a random dial in an agent’s brain, settling her decisions (Pereboom, 2001, pp. 52–3). Similar examples are invoked by Haji (2000, pp. 219–20), and Almeida and Bernstein (2003).

order to tackle the Luck Problem. Levy (2008) criticises this solution as inadequate. He contends that while it might help to show that the agent's reasons explain her choice whichever way she decides, it doesn't explain how the agent's reasons settle the choice itself, fixing how she chooses between the available options:

...avoiding the luck objection requires more than just showing that agents choose for reasons. An adequate libertarianism must offer a reasons explanation of the very choice that is the locus of free will, not simply of the options between which the agent chooses. The process view does not, and cannot, offer such an explanation, and therefore fails to avoid the luck objection. (p. 749)

Levy explicitly notes that this approach may secure *as much* control as the agent would have in a deterministic universe, but notes that that libertarians need to secure *additional* control, i.e., he concedes that this constitutes an adequate response to the Luck Problem, but questions whether it solves the Enhancement Problem. While the Luck Problem rests on the simple mistake of conflating a causally indeterministic process with a causally impotent one, the Enhancement Problem presents a more serious challenge. I agree that appealing to a process view *alone* does not resolve it.

On the picture just sketched, it seems false that the agent's mental life 'disappears' from any explanation of her own choices and action, or that 'nothing about the agent' causally explains her decision. Agents' choices are explained by the way that their decision-making processes unfold. But the Enhancement Problem pushes us to question whether the agent can, in addition, exercise control over the *way* her mental processes unfold. If this is explained by chance, then while her own mental life bears the right sort of explanatory relationship to her final choice, she will still lack any additional power to settle the choice. Unlike the Luck problem, the Enhancement Problem is supported by the *weak* luck claim.

This challenge becomes more pressing when we contemplate why we might suppose that the ability to choose otherwise matters. For leeway incompatibilists, having morally robust alternatives which are accessible in the actual circumstances is important for responsibility. This view is typically supported by concerns about fairness, which seem especially troubling in the case of retributive blame: If an agent is to deserve punishment purely for the sake of retribution (as opposed to forward-directed goals), and if such treatment is to seem fair, then plausibly we must suppose that it was reasonable to expect the agent to have avoided the wrongdoing. It seems unfair to blame an agent who could not possibly have behaved any better in the circumstances.<sup>4</sup>

We hold agents responsible not only for their actual choices, but for choosing in one way as opposed to another. Plausibly then, if alternatives of this sort are to *add* anything to moral responsibility, we had better suppose not only that the agent's mental life is causally effective in fixing her choices, whichever way she chooses, but that features of her mental life causally explain how her choice is settled. We must refute the *weak* luck claim as well as the strong one.

<sup>4</sup> See: Wolf (1980), Otsuka (1998), Moya (2011).

I aim to show how indeterminism may be relevant to our meeting such leeway conditions. I will say little about the opposing source incompatibilist view, according to which what matters is not whether the agent has alternatives, but whether the agent is the ultimate causal source of her choices. It is a separate question whether indeterminism is relevant to that condition.

### 3 The Enhancement Problem

I have argued that an agent's choice will be adequately explained by her decision-making process (and hence by events in her mental life), so long as the way that this process unfolds determines the outcome of her deliberation. But if this it to secure *additional* control, the agent will plausibly also need to exercise control over the way that the process unfolds. How could an agent have this sort of control?

The question essentially amounts to asking how we can exercise agency over the way that we exercise our agency. There seem to be three potential sorts of answer:

- (1) The decision-making process must be fixed by some independent exercise of agency that qualifies it as being within our agential control.
- (2) The decision-making process must be fixed by something external to the process, which is not a further exercise of agency, but which nonetheless grounds its status as being within our agential control.
- (3) There is something intrinsic to the decision-making process itself in virtue of which it counts as being within our agential control.

The first solution leads to regress. If our control over the decision-making process requires some further independent exercise of agency, then we will need to ask how we exercise control over this further exercise of agency, and so on. Since we are not capable of *infinite* exercises of agency (whether an infinite series of earlier decisions or an infinite series of meta-decisions), the regress will have to terminate with something that is not another exercise of agency. If we can only control our agency insofar as it's subject to some additional exercise of agency, freedom becomes conceptually incoherent.

The second approach retains something of the first. It seeks something that might operate as a surrogate for an additional exercise of agency; something that gives some directive or recommendation to guide our decision-making procedure, paralleling the sort of control we exercise over our own actions when we direct them through our decisions. E.g. Perhaps we exercise control over our decisions insofar as they are fixed by higher order volitions (Frankfurt, 1971) or, ultimately, by our deeper values (Watson, 1975, 1996).

Wolf (1990) terms this the 'Real Self' view and criticises it on the basis that agents may not be responsible for their deeper values and may not have been capable of developing better ones. Watson (1996) concedes that this may not suffice for the sort of 'accountability' plausibly required to render retributive blame fair, though

acting in accordance with one's values may suffice for 'attributability' and the latter captures one important sense of responsibility.

However, Wolf herself agrees that there can be a termination to the regress which is not a further exercise of agency. She supposes that we may blame agents only when they were capable of doing better. This involves being able to respond to 'the True and the Good'. If an agent was capable, in the circumstances, of aligning her choices with the True and Good, this might get us the sort of 'accountability' required for retributive blame.

While this second answer does not render freedom incoherent like the first, if these exhaust our options, the Enhancement Problem becomes inescapable.

Whether the regress terminates at our deepest values or at the True and the Good, it is implausible to suppose that these must fix the agent's choices necessarily or uniquely if she is to be responsible. If the agent's choices had to be *necessarily* fixed by the relevant values, then for Watson, an agent could never be responsible for weak-willed behaviour, since it would not cohere with her deepest values, and for Wolf, the agent could never be responsible for bad behaviour, since it would not cohere with the True and the Good.

And if the values had to *uniquely* fix the agent's choices, then presumably agents could never be responsible when choosing between options regarded as equally permissible on the basis of the relevant values.

Instead, it seems we ought to say that if an agent is to be held responsible, it must be possible (rather than necessary) that her choices are suitably *constrained* by (rather than uniquely fixed by) the relevant values.

While this shows that responsibility, on either view, ought to be consistent with alternative possibilities, we are led straight back to our original problem about how agents can settle which alternative is chosen: In the case of Wolf's condition, the agent may be responsible granted that she is *able* to do the right thing for the right reason, even if she does not actually do the right thing, and she may be able to do various things consistent with this constraint. But now we must ask how she can be responsible for actually *exercising* her ability in one way as opposed to another. Similarly, on the Real Self view, we will grant that an agent is responsible insofar as her choices are suitably constrained by her deepest values. But where the agent could choose in more than one way, consistent with these values, we will wonder how she could be responsible for the way this choice is settled.

This gets to the crux of the Enhancement Problem. Either the agent's values must speak decisively in favour of a single choice if she is to be responsible or she could be responsible consistent with resolving the choice in more than one way. In the former case, she can have no alternative. In the latter case, since *ex hypothesi*, the relevant values are consistent with her making either choice, they cannot possibly explain why she chooses one way as opposed to another. If this is the basis for agential control, it is impossible that the agent might exercise such control with respect to settling *which* choice she makes.

It has seemed to many theorists as if these first two options (appealing to a further exercise of agency or appealing some set of base values that might serve as a regress-terminating surrogate for a further exercise of agency) exhaust all of the possibilities. If this were true, the Enhancement Problem would be inescapable. The third option is



rarely explored, perhaps because it seems inherently circular to suppose that an agent's control over the way that she makes decisions could be supplied by anything internal to the decision-making process itself. It has an unnerving feel of bootstrapping about it. I hope to show, however, that it should not be dismissed so readily.

Consider Strawson's response to the problem of induction (1952, ch. 9). Strawson argues that proportioning one's beliefs to the strength of the evidence is what constitutes reasonable assessment of evidence, hence asking whether the inductive method as a whole is reasonable involves asking an ill-formed question, akin to 'is the use of logic logical?' or 'is the law legal?' There may be something similarly suspicious about the question, 'do we exercise agency over the way that we exercise our agency?'

In the case of both logic and legality, the incoherence arises because the concepts pertain to something procedural. An inference is deductively valid insofar as it follows the rules of deductive inference. An action is legal insofar as it accords with rules that result from the right sort of legislative procedure. It is only with reference to the features of proper legal or logical processes that it makes sense to call something legal or logical, and it is conceptually confused to suppose that some independent basis is required to ground the legality and validity of the respective processes themselves.

I want to suggest that what it means to exercise agency with respect to our decisions is for them to be the culminations of a certain sort of procedure. To ask whether the procedure itself is something over which we have agential control is usually to ask an ill-formed question. The Enhancement Problem only seems compelling insofar as we buy into the implicit assumption that we need to exercise agency with respect our agency, or else we need some surrogate of an additional exercise of agency to play a parallel role. Insofar as agential control is regarded as procedural, we needn't buy into this assumption. Instead of searching for some other source of control that determines how we exercise control over our choices, we ought to ask what sorts of process might *constitute* controlled decision-making.

I propose that the following two features are what defines a decision-making process as a genuine exercise of agency suited to settling what to do:

- (1) The process itself could be accurately described as settling which of the outcomes occur.
- (2) The process is an exercise of agential control, where this is analysed in terms of procedural features (and does not require any additional exercise of agential control).

Let's examine these features in turn.

#### **4 When Does a Process 'Settle' an Outcome?**

I have suggested that a process constitutes an exercise of agency when it has the right characteristics, exemplifying a certain sort of procedure. The point, then, is not to appeal to an additional exercise of agency that directs the process, but to

motivate the idea that there are inherent principles which are definitive of agential control over outcomes.

Part of this will involve spelling out what it means for processes (of any sort) to settle outcomes. It is only if something can count both as inherently constituting an exercise of agential control *and* as settling an outcome, that we can start to understand what it means for an exercise of agency to settle an outcome.

Some may suppose that there is something inherently wrong with the notion that any chancy process could count as ‘settling’ an outcome. However, we typically describe many procedures as outcome-settling, and there’s no obvious reason to suppose that they must be deterministic in order to fulfil this role. In fact, a procedure needs in *some* sense to be able to culminate in more than one outcome if it is to be described as settling which outcome occurs. E.g. one could not shoot the prisoners in the head as a method of settling whether or not to execute them. This would simply count as *carrying out* the executions, not settling whether to carry them out.

We frequently talk about outcomes being settled by games, interviews, examinations, competitions, negotiations, etc. Typically, we describe a process as settling an outcome under the following conditions:

- (a) The process is undertaken for the explicit purpose of settling an outcome.
- (b) The process does culminate in one of the outcomes occurring.
- (c) The process could (in some sense) have culminated in a different outcome occurring instead.

While (c) needn’t be read so as to require indeterminism, there is also no obvious reason to think it requires determinism. When we say things like, ‘this match will settle which of these two teams wins the tournament’, it seems doubtful that this would be misleading in some way unless the process should turn out to be fundamentally deterministic.

An indeterministic decision-making process may, it seems, settle an outcome in the same way that other archetypical outcome-settling procedures do. It seems reasonably uncontroversial that decision-making processes meet conditions (b) and (c). They will meet (a) too, if we can make a case for supposing that the *purpose* of decision-making is settling what to do.

Should we suppose that the point of decision-making is settling what to do? Often a contrasting view is presupposed (usually implicitly, sometimes by compatibilists and incompatibilists alike): The view that decision-making is aimed at striving to either to form or to carry out particular intentions. If that view were right, then decision-making would not perfectly parallel other outcome-settling processes. But we have reason to think that view is mistaken.

Perhaps the most famous attempt to sketch out an event-causal picture of indeterministic decision-making (one to which the present discussion owes a lot) is Kane’s (1996, 1999). According to Kane, when an agent makes a choice, she is essentially trying to choose both options, and only one of those attempts can

succeed. It's easy to see how this strategy helps us to tackle the Luck Problem: The mere *possibility* of failure in our attempts is not normally seen as freedom-undermining, where we actually succeed in doing what we were trying to do. It's harder to see how this picture helps with the Enhancement Problem. It seems that removing the *threat* of chance is the best we can hope for.<sup>5</sup> If we can tolerate the possibility of failure in our efforts and still be morally responsible for them, then we can tolerate the possibility of making an alternative choice too. But it's not obvious why we should *require* it.

The problem, I believe, is that we are focused entirely on our first-order goals, and this obscures the real purpose served by making up our minds. At best, it reflects the purpose served by forming particular intentions.

Suppose that Anita is unemployed and desperate for work and has been offered her dream job. It's exactly what she wants to do, it's excellent for her career, well paid, etc. But taking up the position would require a costly sacrifice. She would have to move to another country, and she has some personal ties where she lives which are immensely important to her, and which would be seriously damaged by a move abroad. Now suppose that she is trying to make up her mind about whether to accept or reject the job offer.

It seems wrong to suppose that Anita is both trying to accept the job and trying to reject it. What she is trying to do is to *arrive* at a decision; to make up her mind. We can only feasibly try to do something that we already intend to do. And we cannot form the relevant intention prior to deciding. Until Anita makes up her mind, she does not have an intention to carry out, so trying to carry it out is not an option. As Lemos (2011) notes, we may suppose that agents in such situations *want* to do both things, but we run into difficulties if we suppose that they are *willing* or *trying* to do both.

While there is a significant difference between trying to make a decision and trying to carry out a decision that one has already made, these are surprisingly often conflated. Consider the bizarre worry about how libertarians can make promises (Van Inwagen, 2000; Mele, 2004). The worry is that an agent could not reasonably assure us that she is able to keep a promise if she is also able to decide otherwise; after all, doesn't this simply make it a matter of chance whether she will keep the promise or not?

But having the ability to decide either way does not entail lacking the resolve to carry out one's decision once it has been made. There is all the difference in the world between *making* a choice and *neglecting* one's choice. Possession of an ability to settle what to do does not entail possession of the ability to act with total disregard for what one has settled on doing. Yet, this confusion is common and has a long history. Consider Hobart's claim that if an agent's choice is undetermined, it would be 'an interference, and an utterly uncontrollable interference, with his power of acting as he prefers' (1934, p. 7). Again, this seems to confuse carrying out one's preferred course of action with deciding one's preferred course of action.

<sup>5</sup> See Franklin (2011).

If one confuses settling what to do with carrying out one's intention, this automatically mystifies decision-making. If Anita were trying to accept the job, then, plausibly, the chance of rejecting it anyway would interfere with her goals. The same would be true of the possibility of her accepting the job, if she were intending to reject it. Once she has a settled intention that she is trying to carry out, the possibility of doing otherwise certainly does look like an 'uncontrollable interference' with her ability to pursue her preferences. But if Anita's goal is to make up her mind, this goal is not thwarted by her possessing both the ability to accept the job and the ability to reject it. In fact, she must presuppose that she has both of these abilities if there is to be any *point* in trying to decide between them.

The idea that our goals are structured in a more complex way than we can gauge from considering our first order desires alone is familiar from Frankfurt's (1971) famous discussion. Frankfurt notes that as well as having first order desires, such as wanting to go to the gym or wanting to stay in bed, we also have second order desires—such as a desire not to want to stay in bed. And we may also have second order volitions; preferences about which of our first order desires actually moves us to action. I may actually be moved by a desire to stay in bed, but perhaps I would prefer to be moved by my desire to commit to a morning exercise regime.

Second-order desires and volitions give us meta-level goals; goals regarding our first-order aims and motivations. But they are not the only sorts of meta-level goal that we have. And we can have conflicting goals and values even at a meta-level. Anita may usually wholeheartedly endorse both her desire to have a rewarding career and her desire to sustain significant personal ties. The problem arises when the two are thrown into conflict. Sometimes our goal is to *resolve* such conflicts by settling our priorities. It is one thing to endorse a first-order volition as the one that we want to move us to action; it is another thing to work out *which* of our motivations to endorse when our values come into unexpected conflict.

It is this goal that enables us to see how a decision-making process might count as an outcome-settling process, parallel to other archetypal examples of outcome-settling processes. But it is also with reference to this goal, and how we might practically pursue it, that we can begin to understand the features in virtues of which a decision-making procedure could constitute a controlled exercise of agency. Let's examine this next.

## 5 A Procedural Analysis of Agency

The idea that an indeterministic procedure might settle an outcome, on its own, is not especially mysterious, as we can see when we reflect on archetypal cases, such as exams, negotiations, tournaments, etc. Typically, these are regarded as outcome-settling, at least partly, because that is the explicit purpose for which such activities are undertaken. Deliberation and decision-making, likewise, are undertaken for the express purpose of settling what to do. Much of the confusion surrounding this notion is due to a mistaken picture of decision-making: One that sees it simply as a mechanical process of translating one's already-fixed

preferences into matching intentions (of course, we do sometimes do this, when it's plainly obvious what to do, but such cases barely constitute 'decision-making' at all).

I have argued that decision-making processes can settle outcomes in precisely the same manner that other outcome-settling processes do. The next question we must ask is what sort of agency we could have with respect to the way the process unfolds, and hence with respect to the way that the outcome is settled.

As already noted, no solution is going to be available if we require anything like an additional decision or exercise of agency, via which the agent fixes the way that her decision-making process will go. Rather, any solution to this problem must instead be premised on the idea that there is something *inherent* to the process which defines it as an exercise of agency, where this is not something that has to be grounded in any additional exercise of agency.

Again, this claim parallels Strawson's solution to the problem of induction. Decision-making procedure, in important respects, parallels examples such as legal and deductive procedure. An inference is deductively valid insofar as it accords with the rules of deduction. When we ask what makes deductive logic itself valid, we are asking an ill-formed question. What makes something an exercise of agential control, by a similar token, need not be a further exercise of agential control, but something inherent to the sort of *procedure* that forms the basis of controlled deliberative agency.

I maintain that one's decision-making procedure counts as an exercise of agential control when it is conscious, purposively aimed at resolving the question of what to do, and steered (at least to some degree) by sensitivity to pressures of consistency, so that it's equally aimed at producing coherence within one's values and motivations. Moreover, it seems that when one's values are shaped by procedures of this kind, one's values can be said to have been subject to some degree of critical reflection; this is what it means to live by values that are reflectively embraced, as opposed to values that one holds uncritically.

Let us examine these points in turn. The idea that our decision-making processes are aimed at securing a sort of consistency among our motivations may seem strange. Some theorists put the explanation in the opposite direction: That is, they suppose that already *possessing* consistency among our motivations and values is a prerequisite of rational decision-making, as opposed to something that might be brought about by it.

Frankfurt (1971) argues that we are free to the extent that we wholeheartedly accept the motivations that move us to action. Wholeheartedness requires that we do not suffer ambivalence in relation to our desires. Frankfurt (1992) elaborates on the idea of ambivalence as follows:

The disunity of an ambivalent person's will prevents him from effectively pursuing and satisfactorily attaining his goals. Like conflict within reason, volitional conflict leads to self-betrayal and self-defeat. The trouble is in each case the same: a sort of incoherent greed—trying to have things both ways- which naturally makes it impossible to get anywhere. The flow of volitional or of intellectual activity is interrupted and reversed; movement

in any direction is truncated and turned back. However a person starts out to decide or to think, he finds that he is getting in his own way. (1992, p. 9)

There is, however, an ambiguity about whether harmony among our motivations ought to be regarded as a *prerequisite* of rational decision-making, or as the *goal* of rational decision-making.

We might worry that on Kane's account, on which an agent appears to be trying to do two things at once, there would have to be some degree of ambivalence among the agent's motivations. But I have argued that the point of making up one's mind is not to carry out one's preferences (even if they are in conflict), but to *resolve* a pre-existing conflict and reach a coherent preference-ordering. It is not obvious that this entails ambivalence at all.

If merely being *subject* to a pre-existing conflict of motives (prior to making up one's mind) were regarded as a threat to rationality, then a great many of our choices would count as irrational. It is challenging enough to get one's first-order volitions neatly in line with one's higher-order volitions and deeper values. But the problem runs deeper. Even if we can succeed in getting our own values into some semblance of harmony, the world often presents us with unavoidable dilemmas that *throw* them into conflict.

There is nothing rationally defective about Anita's wanting a rewarding career while also valuing her personal ties. The conflict doesn't arise because of some internal incoherence or greed; it arises only when it transpires that her most promising avenue for pursuing her career goals requires a move abroad which would disrupt her personal ties. Merely being human and tasked with navigating reality is enough to frequently force us into situations where our otherwise coherent values come into conflict. The purpose of deliberation and decision-making is to resolve such conflicts.

While ambivalence is problematic at the point of decision-making, then, just as Frankfurt argues, we cannot expect that harmony among one's motivations must be present *prior* to making a choice. Rather, decision-making is typically aimed at resolving ambivalence and achieving harmony among one's motivations.

Frankfurt's picture of free will involves a state in which one's second order volitions (alongside one's deeper values, presumably) and first order motivations cohere. Let's term this state 'psychic harmony'. Frankfurt's earlier (1971) discussion gives the impression that only an agent whose mind *happens* to be in a state of psychic harmony could have free will. It seems to be a *prerequisite* of rational decision-making. He says little, however, about how one arrives at a state of psychic harmony. Is this something that happens through sheer good fortune? (If our values, second-order volitions, and first-order volitions just *happen* to fall in line, like the spinning panels on a slot machine, we would essentially have won a sort of free will jackpot).

In his later discussion (1992), he explicitly sees deliberation as something that might involve resolving such conflicts. This seems to better cohere with the way we usually think of the decision-making process. There are various higher-order attitudes and aims we may have with respect to our first-order desires and motivations. One of our higher-order goals may be a desire to actively settle our priorities, ordering our values so as to ensure that our commitments cohere. Instead of supposing

that only an agent with a harmonious mind could make rational choices, perhaps we ought to suppose that only an agent capable of rational deliberation and decision-making could *achieve* a harmonious mind.

The purpose of decision-making is not to translate a consistent set of motivations (which we just happen to have) into actions; it's a way of striving to achieve consistency among our motivations, and we do this precisely so that we are able to determine which of our first-order motivations to endorse. This is a deliberate and conscious thought process; a sort of internal effort at conflict resolution. Let's examine precisely what this procedure might involve.

## 6 Internal Conflict Resolution

It's worth reflecting on the fact that there are rational pressures that guide a process of deliberation and decision-making, which are, essentially, forces inherent in the procedure itself, which direct the way that the thought process unfolds. The decision-making process broadly parallels the sort of procedure Rawls (1971) suggests we might use in order to achieve 'reflective equilibrium' between our considered moral judgements and our basic moral principles.

When we are striving to work out what to do in situations of value conflict, we are essentially trying to get our priorities straight; we weigh up different ways of ordering and prioritising our deeper values against the courses of action these would commit us to. The procedure may require us to move back and forth between contemplating our immediate decisions and our deeper commitments, adjusting them so as to create a balance that we can live with, and that will resolve the conflict, speaking decisively in favour of one course of action.

Such a process is not merely one of translating pre-existing preferences into action; it frequently involves *crystallising* our underlying commitments. When we have no conflicts to resolve, we often have no pressing need to straighten out our priorities. When we are not faced with an immediate practical obstacle, we are able to tolerate leaving our priorities vague until a time comes when we are forced to subject them to deeper scrutiny. Situations of motivational conflict may force us to clarify our values; the immediate practical challenge of working out what to do may prompt us to deal with the broader challenge of working out which commitments we are prepared to live our lives by.

Until Anita receives her job offer, she may never have to weigh up whether her personal ties or her career goals mean more to her. But she can hardly avoid settling this further question, about what she ultimately values, when she is faced with the practical problem of deciding whether or not to take the job. She cannot solve the problem of whether to accept the job without also determining whether her career goals mean more to her than the personal ties that she would have to sacrifice in order to pursue them.

There are a number of reasons why we might think that such a procedure is definitive of agential control. Firstly, it involves conscious thought. This contrasts sharply with cases in which we act thoughtlessly, or cases in which we are driven by unconscious drives, such as implicit biases that we may consciously repudiate. It

also contrasts with cases that bypass our thought processes, such as cases of external manipulation.

Moreover, the process itself is steered (albeit to varying degrees) by sensitivities to certain sorts of rational pressure; the rational pressure to have a consistent structure of values and motivations, which frees us from ambivalence and internal conflict. It reflects our drive to have motivations and reflective judgements that cohere. This contrasts sharply with cases in which we are driven by motives that are *not* sensitive to any urge to resolve conflicts with our other values and judgements—such as typical cases of irrational phobia, compulsion, or addiction—or choices that are entirely ‘wanton’ in Frankfurt’s sense.<sup>6</sup>

Finally, it captures what it means to have values that are shaped by some degree of critical reflection, as opposed to merely being there through indoctrination or unreflective habit. The procedure not only involves being passively subject to values and then translating them mechanically into matching intentions. Rather, it involves working out which values one is prepared to live by through conscious reflection.

My claim is that it is these intrinsic characteristics which make a decision-making procedure *count* as control-conferring (being a conscious, purposive, consistency-steered rational thought process). The procedure is control-conferring precisely because it is undertaken for a conscious purpose, and precisely because it unfolds in accordance with the sorts of rational principles that make it, essentially, a way of finding resolutions to the sorts of internal conflicts that impinge psychic harmony. The demand for a source of control located outside of the procedure itself and independent of the intrinsic features that make the procedure control-conferring is ill-formed.

Note that the view that there may be something about the process that *defines* it as an exercise of controlled agency importantly differs from the view that the process is controlled in virtue of being *directed* by an exercise of controlled agency. The latter gives rise to a puzzling regress, which the former avoids.

On this picture, deliberation is not within our control on the basis that we stand outside of the process, causally influencing it. Rather, it is the fact that the process constitutes the procedure via which an agent consciously seeks to coherently structure and settle her values that grounds the sense in which it is connected to the self. What constitutes our deeper values, or our ‘true self’ typically *depends* on this sort of this procedure. It is via these procedures that we commit to certain values and priorities as our own, and it is this process of so committing that confers a sense of control.

---

<sup>6</sup> One complication is that some choices are irrational, but not ‘wanton’ or compulsive. While some *degree* of sensitivity to rational pressures seems to be required for responsibility, it seems we ought to tolerate *some* degree of irrationality too. It’s far from clear precisely where we draw this line.



## 7 Solving the Enhancement Problem

We are finally in a better position to assess the Enhancement Problem. Proponents concede that the right sort of process may suffice for moral responsibility, and that that the important features of such a procedure may be present *consistent* with indeterminism. But they object that there is nothing *extra* that indeterminism could add that might be relevant to moral responsibility.

While there are theorists who are sceptical about the moral relevance of the ability to choose otherwise (Frankfurt, 1969), the Enhancement Problem rests, instead, on the idea that nothing could *constitute* the relevant ability. This may be motivated, as we saw, by the weak luck claim. The idea is that even if an agent's process of decision-making could have gone differently, the *agent* could never have the ability to control whether it goes differently or not. Rather, this would simply be settled by luck, and thus the agent could have no power with respect to settling how she chooses.

Someone with *this* worry may concede that the process I describe is itself an exercise of agential control, but will demand that we explain how the agent can, in addition, exert control over how the process itself unfolds. I have argued that this is an illegitimate demand. The process via which the agent exerts control is not the sort of thing over which she needs to exert some *additional* form of control. Such control does not fundamentally depend on independent exercises of agency. Rather, agential control is defined in terms of the inherent features of the process itself.

I do not mean to rule out the *possibility* that an agent sometimes might succeed in exercising additional agency with respect to which future choices she makes. E.g. the agent might decide that she would like to be the sort of person who chooses what is right over what is easy, and might take positive steps to develop her character so that she chooses what is right more often in future.<sup>7</sup> But I do want to deny that additional exercises of agency are *required* in order for any given exercise of agency to count as within the agent's control. If this were always required, we would end up with a regress, which would inevitably terminate with something that is *not* an exercise of agency. My claim is that an agent may well have control over her own agency in virtue of its internal structure alone.

In fact, when we say that an agent can settle which outcome occurs, all we could intelligibly *mean* is that the outcome is settled by the sort of process that constitutes an exercise of agential control. We have, then, a rival analysis of what it would mean for an agent to settle an outcome, and we can compare it with the sort of sort of analysis that underlies the Enhancement Problem:

*The Additional Agency Analysis (AA):*

An agent is able to settle which decision she makes only insofar as (a) her decision-making process settles which decision she makes, and (b) she is capa-

<sup>7</sup> I am grateful to an anonymous reviewer for bringing this point to my attention.

ble of exercising an independent sort of agential control over the way that her decision-making process unfolds.

*The Procedural Analysis: (PA):*

An agent is able to settle which decision she makes only insofar as (a) her decision-making process settles which decision she makes, and (b) her decision-making process inherently constitutes an exercise of agential control.

Which of these provides a more satisfying analysis? No doubt some theorists will maintain that nothing short of AA will do, and that PA is a poor substitute. This is likely to be the view of someone who thinks that AA is what full-blown control would require (anyone who supposes, like Galen Strawson (1994), that freedom requires a Godlike ability to be entirely self-made). But this stance is rendered implausible by the fact that the demand in question, upon further scrutiny, seems ill-formed. We can sensibly ask what the definitive features are of the sort of procedure that constitutes an exercise of agential control, but it is rarely intelligible to ask whether we exercise agential control over the way that we exercise agential control.

I maintain that purely retributive blame would only be fair if the agent had a reasonable opportunity to avoid blame in the circumstances, and I think this is rendered doubtful if it's literally *impossible* for the agent to avoid blame given the way those circumstances actually are (including the way that the past and the laws of nature are). But nor does it seem to me that the only alternative is to suppose that the agent would need to have abilities that are magical and Godlike if the expectation that she should avoid blame is to become any more reasonable.

Rather, it seems there is a good case for supposing that an agent would have a fair opportunity to avoid blame (and that it would not be unreasonable to *expect* the agent to avoid blame), if the agent were able to avoid blame in a sense analysed as meaning:

- (a) That she could have decided otherwise given the actual circumstances, and
- (b) She was able to settle how she decides in precisely the sense specified by PA; whereby her own conscious purposive thought process is both what constitutes her exercise of agential control and what settles her decision.

Compare the following excuses:

*The hard determinist's excuse:*

I admit, I did something immoral, and my choice was the result of a conscious, consistency-steered, rational process of deciding what to do. However, I could not possibly have avoided blame in the actual circumstances. I made the best choice I possibly could consistent with the laws of nature and the way things were in the past.

*The further-agency excuse:*

I admit, I did something immoral, and my choice was the result of a conscious, consistency-steered, rational process of deciding what to do. Moreover,

I admit that it was possible for me to avoid blame in the circumstances. However, I deny that I was able to avoid blame as my immoral action was merely explained by the way my deliberate exercise of agency went. I can't exercise any further agency with respect to how my agency goes.

It makes sense to think that indeterminism *adds* something relevant to responsibility if we are inclined to suppose that the further-agency excuse is weaker than the hard determinist's excuse. There will be some who think that any expectation that the agent ought to have avoided blame is unreasonable in both cases and others who think it is reasonable in both. What I hope to have shown, however, is that it's far from obvious that the conditions I outline have no bearing on the fairness of our expectations at all.

There is at least a case to be made for supposing that if an agent has the ability to choose otherwise alongside an ability to settle how she chooses in the sense captured by PA, this might be relevant to the question of whether it was reasonable to expect her to choose otherwise.

Doubt about the relevance of the ability to choose otherwise comes almost entirely from the sense that nothing could enable us to make sense of this ability. This article has aimed to show that we can make sense of it. This ability will be incoherent only insofar as we think that any exercise of controlled agency must be predicated on something akin to a further exercise of agency. But this assumption can be questioned. Moving towards a procedural view of control enables us to see it in terms that do not give rise to a regress and that demystify the notion that an agent might be able to settle her decisions.

**Acknowledgements** I am grateful to Dr Tuomas K. Pernu, Professor John Lemos, and to anonymous reviewers for valuable comments and discussion on this article.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Almeida, M., & Bernstein, M. (2003). Lucky libertarianism. *Philosophical Studies*, 113(2), 93–119.
- Collins, A. (1717). *A philosophical inquiry concerning human liberty*. Golden Lion in St Paul's Church Yard.
- Elzein, N. (2019) The demand for contrastive explanations. *Philosophical Studies*, 176, 1325–1339.
- Fenton-Glynn, L. (2017). A Proposed probabilistic extension of the Halpern and Pearl definition of 'actual cause'. *The British Journal for the Philosophy of Science*, 68(4), 1061–1124.
- Frankfurt, H. G. (1969). Alternate possibilities and moral responsibility. *The Journal of Philosophy*, 66(23), 829–839.
- Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy*, 68(1), 5–20.

- Frankfurt, H. G. (1992). The faintest passion. *Proceedings and Addresses of the American Philosophical Association*, 66(3), 5–16.
- Franklin, C. E. (2010). The problem of enhanced control. *Australasian Journal of Philosophy*, 89(4), 687–706.
- Franklin, C. E. (2011a). Farewell to the luck (and Mind) argument. *Philosophical Studies*, 156(2), 199–230.
- Haji, I. (2000). Indeterminism, explanation, and luck. *The Journal of Ethics*, 4(3), 211–235.
- Haji, I. (2001). Control conundrums: Modest libertarianism, responsibility, and explanation. *Pacific Philosophical Quarterly*, 82, 178–200.
- Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy*, 98(6), 273–299.
- Hobart, R. E. (1934). Free will as involving determinism and inconceivable without it. *Mind*, 43, 1–27.
- Hume, D. (2000). *A treatise of human nature*. Oxford University Press.
- Kane, R. (1996). *The significance of free will*. Oxford University Press.
- Kane, R. (1999a). Responsibility, luck and chance: Reflections on free will and indeterminism. *Journal of Philosophy*, 96, 217–240.
- Lemos, J. (2011). Wanting, willing, trying and Kane's theory of free will. *Dialectica*, 65, 31–48.
- Lemos, J. (2018). *A pragmatic approach to libertarian free will*. Routledge.
- Levy, N. (2008). Bad luck once again. *Philosophy and Phenomenological Research*, 77(3), 749–754.
- Lewis, D. K. (2000). Causation as influence. *The Journal of Philosophy*, 97(4), 182–197.
- McCall, S., & Lowe, E. J. (2005). Indeterminist free will. *Philosophy and Phenomenological Research*, 70, 681–690.
- Mele, A. R. (1999a). Ultimate responsibility and dumb luck. *Social Philosophy and Policy*, 16(2), 274–293.
- Mele, A. R. (1999b). Kane, luck, and the significance of free will. *Philosophical Explorations*, 2(2), 96–104.
- Mele, A. R. (2004). Can libertarians make promises? *Royal Institute of Philosophy Supplement*, 55, 217–241.
- Mele, A. R. (2005). Libertarianism, luck, and control. *Pacific Philosophical Quarterly*, 86(3), 381–407.
- Mele, A. R. (2006). *Free will and luck*. Oxford University Press.
- Moya, C. (2011). On the very idea of a robust alternative. *Critica*, 43(128), 3–26.
- Otsuka, M. (1998). Incompatibilism and the avoidability of blame. *Ethics*, 108, 685–701.
- Paul, L. A. (2000). Aspect causation. *The Journal of Philosophy*, 97(4), 235–256.
- Pereboom, D. (2001). *Living without free will*. Cambridge University Press.
- Pereboom, D. (2004b). Is our conception of agent causation coherent? *Philosophical Topics*, 32, 275–286.
- Pereboom, D. (2014). *Free will, agency, and meaning in life*. Oxford University Press.
- Rawls J. (1971) *A theory of justice*. Second Edition, 1999, Harvard University Press.
- Smart, J. J. C. (1961). Free will, praise and blame. *Mind*, 70(279), 291–306.
- Strawson, G. (1994). The impossibility of moral responsibility. *Philosophical Studies*, 75(1), 5–24.
- Strawson, P. F. (1952). *Introduction to logical theory*. Methuen.
- Van Inwagen, P. (1983). *An essay on free will*. Oxford University Press.
- Van Inwagen, P. (2000). Free will remains a Mystery: the eighth philosophical perspectives lecture. *Philosophical Perspectives*, 14, 1–19.
- Waller, B. N. (1990). *Freedom without responsibility*. Temple University Press.
- Watson, G. (1975). Free agency. *The Journal of Philosophy*, 72(8), 205–220.
- Watson, G. (1996). Two faces of responsibility. *Philosophical Topics*, 24(2), 227–248.
- Wolf, S. (1980). Asymmetrical freedom. *The Journal of Philosophy*, 77(3), 151–166.
- Wolf, S. (1990). *Freedom within reason*. Oxford University Press.
- Woodward, J. (2003). *Making things happen: a theory of causal explanation*. Oxford University Press.
- Woodward, J., & Hitchcock, C. (2003). Explanatory generalizations. Part i: a counterfactual account. *Noûs*, 37, 1–24.