

A Modest Classical Compatibilism

Matthew J. Hart
University of Liverpool

BIBLID [0873-626X (2017) 45; pp. 265–285]

DOI: 10.1515/disp-2017-0007

Abstract

The advent of Frankfurt-style counterexamples in the early 1970s posed a problem not merely for incompatibilists, but for compatibilists also. At that time compatibilists too were concerned to hold that the presence of alternative possibilities was necessary for moral responsibility. Such a classical compatibilism, I argue in this paper, should not have been left behind. I propose that we can use a Kratzer-style semantics of ‘can’ to model ‘could have done otherwise’ statements in such a way that the truth of such expressions is both (i) evidently consistent with determinism, and (ii) clearly such that Frankfurt-style counterexamples do not count as cases where the agent could not have done otherwise.

Keywords

Classical compatibilism; Angelika Kratzer; alternative possibilities; moral responsibility; Frankfurt-style counterexamples.

1 Introduction

The Principle of Alternate Possibilities (PAP) has been an important principle in the debate concerning free will and determinism. Here is a statement of it drawn from Widerker and McKenna (2006: 2).

PAP: A person is morally responsible for what he has done only if he could have done otherwise.

The principle states a necessary condition for acting in a morally responsible fashion—no alternative possibility, no moral responsibility. But why believe PAP? I see it as a presupposition of a certain conversational practice of ours. Often, when an individual is accused of some sort of misdemeanor and he desires to be exonerated from the accusation, he will protest using language of the following sort:

Disputatio, Vol. IX, No. 45, October 2017

Received: 27/07/2017 Accepted: 02/11/2017

© 2017 Matthew J. Hart. Creative Commons Attribution-NonCommercial-NoDerivs 3.0 License

‘I couldn’t help it!’ or ‘But there was nothing I could have done!’ or ‘I wasn’t able to do anything about it!’. Moreover, it isn’t common to hear a retort of the sort, ‘It doesn’t matter that you couldn’t help it—you are still guilty’. A response like that merely baffles. The far more common way to convince the man of his guilt is to say, ‘No, you could have helped it. You could have done such-and-such.’ We can see PAP, therefore, as a codification of the principle that governs this practice. We therefore appear to accept that disclamations of the ability to do otherwise (if true) are sufficient for exculpation. But if the absence of this ability is sufficient for the absence of moral responsibility, then the presence of the ability must be necessary for moral responsibility. Hence, PAP.

PAP is an important premise in what I call the Naive Ability Argument for Incompatibilism. The argument is as follows:

- (1) If someone is morally responsible for performing an action at t , then they could have done otherwise at t .¹
- (2) If determinism is true, then, for any time t , no-one could have done otherwise than they did at t .²

Ergo,

- (3) If determinism is true, then no-one is morally responsible for performing any action.

(1) is an expression of PAP, and (2) is the intuitive suggestion that if a description of the past up till t in conjunction with the laws of nature entails that you perform some action A at t , then you could not, at t , have done otherwise than A.

¹ Advocates of PAP admit that you may be morally responsible in cases of *derivative* responsibility for an action you performed at t even though you couldn’t have done otherwise at t . For example, you are usually still responsible for the things you can’t help doing when drunk because you usually could have refrained from getting drunk in the first place. But I will take it as read that it is *non-derivative* moral responsibility that this paper is concerned with.

² The first instant of time, if there is such, may be considered an exception to this. But since human beings and their acts came to be after the first instant had been and gone, this point is immaterial. Make t a variable ranging over all instants later than the first instant, if you wish. See Bailey 2012 for discussion.

Compatibilists are split over how to respond to this argument. Historically, compatibilists denied premise (2); until the 1970s compatibilists were keen to construe abilities in such a way that it was clear that, if their account were correct, then determinism didn't remove the ability to do otherwise. Such compatibilists are now called 'classical compatibilists' (see Berofsky 2006). They typically defended conditional analyses of ability, saying something like the following:

(CAB) An agent *S* has the ability to do otherwise at *t* iff were *S* to try/sufficiently desire/intend to do otherwise at *t*, they would do otherwise.

How they construed the content of the antecedent varied from one such compatibilist to another, but it is clear that the truth of determinism is not itself any obstacle to our possessing abilities understood in this fashion. The truth of determinism entails that they *won't* try/sufficiently desire/intend to do otherwise; nevertheless, *were* this to happen, they would do otherwise.

But compatibilists fell away from this view. Beginning in the 1970s, a new breed of compatibilist appeared: the semi-compatibilist. With John Martin Fischer as their leader, they grew in numbers and influence until they became the dominant strain in compatibilist thought. The semi-compatibilist responds to the argument by denying (1)—he judges it a mistake to suppose that the ability to do otherwise is necessary for moral responsibility.

What accounts for this changing of the guard? What gave the semi-compatibilist the *chutzpah* necessary to deny such an intuitive principle? Several things. On the one hand, there were Austin (1961) and Lehrer's (1968) attacks on CAB. But the chief catalyst of change came in 1969 with Frankfurt's influential paper 'Alternate Possibilities and Moral Responsibility'. In that paper Frankfurt offered several counterexamples to PAP; a developed version of the most celebrated of them is as follows.

Black is a neuroscientist of considerable expertise who hates Smith. Indeed, Black hates Smith so much that he desires Smith dead. Now Black also knows that Jones hates Smith. One day Black finds out, to his delight, that Jones has formed a plan to

murder Smith. But Jones is a temperamental fellow, and Black is worried that Jones might change his mind or that his nerve might fail him. So Black implants a device in Jones's brain without Jones's knowledge. This device monitors Jones's brain activity, and as soon as there is any indication that Jones is not going to follow through on his plan, then Black will use the device to cause Jones to kill Smith. As it happens, Jones shows no sign of renegeing on his plan, and he murders Smith.

Surely it is true, says Frankfurt, that (i) Jones is morally responsible for killing Smith and that (ii) Jones was not, thanks to Black's device, able to do otherwise. So, PAP, says Frankfurt, is false. Semi-compatibilists agree.

After word of these things spread, classical compatibilism fell out of fashion and the semi-compatibilist approach became more popular. Frankfurt changed everything, or at least the face of compatibilism. But I think that was a mistake. I shall defend here a modest classical compatibilism that assumes that it is not, strictly speaking, *abilities* that PAP is concerned with, but a certain species of *possibility*. My account builds on the Lewis-Kratzer understanding of 'can', and it permits us another way of seeing that the truth of determinism is compatible with the possibility of doing otherwise. In the next section, section 2, I begin by explaining the semantics Angelika Kratzer developed for understanding modals like 'could' and 'can'. In section 3, I clear away some ambiguity in the expression under discussion: 'S could have done otherwise'. Section 4 contains the application of Kratzer's apparatus to that same expression. In section 5, I show how my account is able to neutralise Frankfurt's counterexample, and also deal with some other concerns.

2 Analysis of modals

David Lewis, in a famous 1976 paper on the paradoxes of time travel, considers whether we ought to say, of a man who has travelled back in time and who is appropriately situated to kill his grandfather, that he *can* kill his grandfather. On the one hand, we can suppose he satisfies all the ordinary criteria for being able to do it (well-armed, fit, has an excellent opportunity), and therefore that he can; but on

the other hand, it appears sound reasoning that he can't, for we can infer from the man's very existence that his grandfather survives and reproduces. Lewis proposes the following resolution: 'To say that something can happen means that its happening is compossible with certain facts. Which facts? That is determined [...] by context. [...] What I can do, relative to one set of facts, I cannot do, relative to another, more inclusive, set.' (1976: 150). So, relative to the fellow's intrinsic properties and opportunity, he can kill his grandfather, but relative to a broader set of facts that includes his grandfather's future existence, he cannot. Disappointingly, beyond the resolution of this paradox, Lewis doesn't go on to deploy this framework as part of a general theory of ability ascriptions.

But Kratzer does. Her approach, like Lewis's, sees context as supplying a set of facts relative to which the 'can' claim is assessed. In her influential 1977 and 1981 papers (reprinted and revised in her 2012a collection), she offers a premise semantics for the modals 'must' and 'can', according to which 'must' functions to express the logical consequence of a proposition from a set of premises and 'can' functions to express the compatibility of a proposition with a set of premises.

By way of illustration, she asks us to consider the following expressions:

- (A) All Maori children *must* learn the names of their ancestors.
- (B) The ancestors of the Maoris *must* have arrived from Tahiti.
- (C) When Kahukura-nui died, the people of Kahungunu said:
Rakaipaka *must* be our chief.

The three expressions all contain a 'must' but the 'musts' appear to belong to different modalities. The 'must' of (A) appears deontic and so concerns something like duty; the 'must' of (B) looks like it is epistemic; and the 'must' of (C) looks like it concerns something like prudence: what would be best for the people of Kahungunu.

So, given the apparently different foci of these 'musts', what is their common element? Kratzer thinks that sentences (A)–(C) are incomplete. To make them complete we add something like the following:

(A') *Given the duties incumbent on Maori children*, they must learn the names of their ancestors.

(B') *Given what we know*, the ancestors of the Maoris must have arrived from Tahiti.

(C') When Kahukura-nui died, the people of Kahungunu said: *In view of what is best for the tribe*, Rakaipaka must be our chief.

The italicised portions indicate the clauses that were suppressed in the original expressions, what Kratzer calls the *conversational background*. It is these hidden portions that function to pick out the set of premises that the central proposition asserted is assessed relative to. Following Kratzer's (2012b) terminology, let's call this picked out set of premises the 'modal restriction'. So we can distinguish three elements in the resulting picture: the modal force—a 'must' or a 'can'; the modal scope—the central proposition explicitly asserted; and the modal restriction—a contextually supplied set of premises relative to which the modal scope is assessed.

If we say all that, then we can give an account of the meaning of 'must' and 'can' that isn't a long and tiresome disjunctive one that makes reference to every different sort of modality. Instead we can just say this:

'Must' expresses that the

<modal restriction> logically implies the <modal scope>

and 'can', being the dual of 'must', expresses that the

<modal restriction> is logically compatible with the <modal scope>

It isn't hard to see how we get the modal scope—it is the central proposition asserted—but what of the modal restriction? How do we arrive at that? The answer lies in the conversational backgrounds. In Kratzer's view these are contextually provided functions that assign to every possible world a set of propositions (the premises). By way of example, consider (B') above. The function that context provides is an epistemic one, concerned with what a certain group of people know. So for each possible world the function assigns a set

of propositions (or premises) that include all and only that which is common knowledge at that time for that community in that world.

So with f as the contextually supplied function from the set of all possible worlds, and w and p being variables ranging over worlds and propositions respectively, we can put Kratzer's central idea in symbols (where the ' \Rightarrow ' denotes strict implication).

'Must' expresses that

$$f(w) \Rightarrow p$$

and 'can' expresses that

$$\sim(f(w) \Rightarrow \sim p).$$

This much more or less captures Kratzer's early 1977 work on the topic.³ But her later 1981 work gave a more complicated picture. She later suggested that the conversational background is really composed of two functions from worlds to sets of premises, namely a *modal base* f and an *ordering source* g . The modal base functions, as above, to get us a set of premises, and the ordering source does the same. But the set of propositions derived by the ordering source is used to induce an order on the worlds that the modal base is true at. Then 'must p ' holds iff p is true in all worlds closest to the ideal specified by the ordering source, and 'can p ' holds iff p is true in at least one of the worlds closest to the ordering source.

I do not wish to go into detail about this later account and the reasons for introducing this more complicated apparatus. I will, for ease of understanding and accessibility, work with the framework given by Kratzer's earlier account. However, for those familiar with Kratzer's later work, I will make it plain in a later footnote how my theory should be stated relative to her more developed account.

3 Analysis of 'could have done otherwise'

We now need to apply all this to PAP. PAP is the claim that the truth of the expression 'S could have done otherwise' is necessary for S

³ I have left out how she handles inconsistent modal restrictions, but that is immaterial for my purposes because, on the account I will give, it will not be possible for the modal restriction to be inconsistent.

to be morally responsible for whatever act S performed. But ‘could have done otherwise’ is an ambiguous expression in at least two ways. Firstly, ‘could have’ is ambiguous between a subjunctive and an indicative reading.⁴ On the indicative reading it occurs as the past form of ‘can’, where ‘can’ is understood to mean ‘be able to’, such that ‘S could have done otherwise’ means ‘S was able to do otherwise’.⁵ On the subjunctive reading, ‘could have’ is the past form of the present tense subjunctive ‘could’. This subjunctive interpretation doesn’t have abilities in view, and instead holds that it is various circumstances that might have obtained that are in view, various possibilities that could have come to pass. On such a reading, the ‘could have’ can be viewed as equivalent to the non-epistemic ‘might have’, such that ‘S could have ϕ -ed’ means ‘S might have ϕ -ed’, rather than ‘S was able to ϕ ’.

So, the indicative ‘could have’ deals with ability, and the subjunctive ‘could have’ talks about possibility. But once we are aware of this, the following idea might occur to one: rather than assuming that PAP is concerned with the *ability* to do otherwise, we might instead suppose that it is concerned with the *possibility* of doing otherwise. We could assume that it is only the truth of the subjunctive ‘could have’ that is necessary for moral responsibility, while supposing that the truth of the indicative ‘could have’ is not, strictly speaking, necessary. In this paper, I shall make this assumption. I say: grant me this modest assumption, and I can give you a classical compatibilism that evades the standard objections.

A great advantage of making this move is that one avoids entirely the debate about satisfactory analyses of ability ascriptions. Abilities aren’t in view at all, only a certain sort of possibility.

But is it a plausible move to make? Someone might make the following objection: ‘But we don’t talk only about whether or not someone could have done otherwise, we also talk about, in present

⁴ A distinction van Inwagen ably deploys in his 1984 response to Dennett 1984.

⁵ Or perhaps, more carefully, ‘S was both able and had an opportunity to do otherwise’, if we think that the ‘have’ modifier forces our attention to a particular occasion. ‘I could run a marathon’, where ‘could’ is the past form of the abilitative ‘can’, talks about a general ability possessed in the past; ‘I could have run a marathon’ suggests that one was able and also in a good position to run a marathon on a particular occasion. I thank Simon Kittle for this point.

cases, whether some individual *can* do otherwise, and ‘can’ is in the indicative.’ It is true that ‘can’ takes the indicative, but it would be a mistake to think that, because of this, it must always be abilities rather than possibilities in view. A distinction is often drawn between the “can” of ability’ and the “can” of possibility’ (see Vetter 2015: 76). That there is a clear possibilist use of ‘can’ can be seen from the following examples: ‘You got burgled? Well, these things can happen.’ ‘Learning a language can be a difficult affair.’ ‘Propositions can be true; contradictions cannot be true.’ They all resist a parsing into ability language. ‘Burglaries are able to happen.’ ‘Learning a language is able to be a difficult affair.’ ‘Propositions are able to be true, but contradictions aren’t able.’ All such expressions, even if one might not be prepared to say that they are strictly incorrect, are nevertheless decidedly awkward and unnatural. This is because we are trying to get ability-talk to do a job better done by possibility-talk. One way of detecting the ‘can’ of possibility is to see whether the sentence carries an identical meaning if you replace the ‘can’ with ‘may’ (just so long as it is clear that the ‘may’ is not expressive of permission or the deontic modality). This works for the above examples. ‘Burglaries may happen.’ ‘Learning a language may be a difficult affair.’ ‘Propositions may be true, but contradictions may not.’ This is because ‘may’ can also be used to express possibility.

So, I can accommodate currently occurring moral-responsibility contexts by supposing that the ‘can’ in the ‘can do otherwise’ that occurs on such occasions is the ‘can’ of possibility. It might be suspected that I press for possibility rather than ability because I think that Kratzer’s apparatus, for all its success, cannot satisfactorily handle abilities.⁶ I agree that Kratzer’s account appears in its best light when it is seen as giving a model of the ‘can’ of possibility, rather than the ‘can’ of ability. Nevertheless, I *do* think that the Kratzer semantics can give an adequate analysis of ability ascriptions, but I want to save all that for another occasion. Defending that supposition would involve, among other things, responding at length to the formal objections made against the idea by Anthony Kenny (1975: 136–7).

⁶ Maier’s pessimistic judgement that ‘the Kratzer semantics alone does not suffice to settle questions about the agentive modalities.’ (2013: 115) appears to be a popular one.

It might also be objected, however, that we also use ‘able’-talk when we talk about whether or not someone ‘can do’ or ‘could have done’ otherwise. We also ask ‘are they *able* to do otherwise?’ and ‘were they *able* to do otherwise?’. We do indeed use such expressions, but, on the view I am proposing here, this is a mistake, though a very understandable one. Given the incredibly close connection between the ‘can’ of ability and the ‘can’ of possibility, and the subjunctive ‘could have’ and the indicative ‘could have’, we can’t expect the ordinary-language user to pay careful attention to such subtleties, especially in a context that concerns agents. Indeed, we would positively expect him to slide, in his benighted ignorance, from possibility-talk to ability-talk and back again, thus confounding what should be kept separate.

One last objection that might be made at this juncture is this: if I am eschewing abilities as, strictly speaking, unnecessary for moral responsibility, then it might be wondered if the account I am giving is really classically compatibilist at all. Weren’t the classical compatibilists of old all concerned to defend the relevance of ability ascriptions to moral responsibility? Well, I agree that they were, but it is also nevertheless clear that I am a classical compatibilist of some sort. I accept PAP, and therefore accept that the truth of a ‘could have done otherwise’ expression is necessary for moral responsibility. Moreover, I don’t say anything like, ‘People naively supposed that alternative possibilities were necessary for moral responsibility. But that was a mistake. What we are *really* interested in when we ask whether or not someone could have done otherwise is the following feature ...’. That is the sort of thing semi-compatibilists say. I am not to be counted amongst their number. I accept the requirement for alternative possibilities. That said, I recognise that the account I am giving is not as fully fledged or as substantial as other classical compatibilisms—but that is why I call it ‘modest’.

Finally, let’s return to the second way in which ‘could have done otherwise’ is ambiguous: we don’t know what doing otherwise involves. I shall take it to mean intending otherwise. I think examples like Fischer and Ravizza’s ‘Sharks’ case show this (1998: 125). John is walking by a pool. He sees a child drowning. He is not inclined to get himself wet, so he continues walking. We think John is a wicked man, and we are inclined to blame him for not saving the child.

However, unbeknownst to John, there were sharks in the water, and, were John to have attempted to rescue the child, he would have been set upon and eaten. So, he couldn't have saved the child in any case. Does this mean we cannot blame him? After all, he couldn't have done otherwise than fail to save the child. But I think that only shows that he must be guilty on account of the presence of a different alternative possibility, namely, that he could have *decided* to try and save the child yet did not. But such decidings and choosings are, I shall suppose, simply the forming of intentions. So, we can hold John guilty for not *intending* otherwise than he did.

4 Applying the Kratzer semantics to 'could have done otherwise'

So, if we are understanding 'could' in the way proposed by Kratzer's early model, then we need to ask ourselves: what is the modal restriction? It is clear what the modal scope is: it is 'S intends otherwise at t '. But if 'could' expresses the compossibility of a set of premises with the modal scope, then we need to know what these premises are. What sort of conversational background does our 'could have done otherwise'-talk presuppose? If we use ' f ' as a variable ranging over functions that perform the role of the modal restriction, then we can use ' f_m ' as a variable ranging over a certain subset of those functions, which subset handles the sort of moral-responsibility contexts we are concerned with. f_m , I shall suppose, will vary according to which particular occasion of an agent's acting is being considered. For every possible world in which the agent in question acts on that one particular occasion in question, f_m will assign to that world a non-empty set of propositions. f_m is therefore a function from worlds to sets of premises; but what are these premises to which it takes us?

If we are to cast the incompatibilist in a Kratzerian mould, then he might say this: f_m takes from every possible world two things in forming the premise set; it takes (i) the laws of nature that hold at that world (N) and (ii) a complete description of the history of the universe of that world that includes all and only such moments that are prior to the agent performing his act at t (D).⁷ And, indeed, if it is true that

⁷ Though an incompatibilist such as van Inwagen (1984) would, if I read him correctly, take the question of the compossibility of N and D with doing other-

(4) $\{N, D\} \Rightarrow \sim(S \text{ intends otherwise at } t)$,

it therefore follows, if the incompatibilist is right, that S could not have intended otherwise relative to this set of premises.

But we don't need to take the incompatibilist's word for it that this is how to characterise the function in question. It is my contention that we should take f_m to be a function from worlds to propositions describing psychological setups such that the relevant possibility of doing otherwise consists in something like *psychological compossibility*—the idea that S's intending otherwise at t must be compossible with S's psychological setup prior to t in order for the morally relevant alternative possibility to obtain. What do I mean to include in these psychological setups? Let me explain.

I introduce my account by way of the following distinctions made by Robert Kane concerning the ambiguity of the term 'will'. Kane says we ought to distinguish the following three senses of the word 'will':

- (i) what I *want, desire, or prefer* to do
- (ii) what I *choose, decide, or intend* to do
- (iii) what I *try, endeavor or make an effort* to do (Kane 1998: 26)⁸

He calls (i) the desiderative or appetitive will, (ii) the rational will, and (iii) the striving will (1998: 27). The desiderative will is easy to understand: it is your desires, your wants. The rational will concerns the decisions you make, typically (at least in part) on the basis of your desires; if you like to drink coffee, you may decide to purchase some. The distinction between the rational and the striving will is more subtle, but it is recognised once we realise that deciding to buy some coffee isn't sufficient for striving for some coffee. I may decide to buy some coffee while I am at work, but because I know that I won't have the opportunity to buy some coffee until I leave work, I only try or

wise as a *consequence* of his account of ability, rather than as an analysis of it.

⁸ To Kane's list we might also add commands as a type of will. To flout the commands of a monarch is to flout his will, even though the monarch might not desire or intend, for some reason(s), that his commands be fulfilled.

endeavour to buy some coffee after I leave the office.⁹

Kane also says, ‘If there is indeterminacy in free will, on my view, it must come somewhere between the input and the output—between desiderative and rational will.’ (1998: 27). And this, I think, is right. I think all parties to the debate, whether compatibilist, libertarian or whatever, must acknowledge this threefold distinction of will. Moreover, we all also realise that the locus of the free-will debate is found here between the desiderative will on the one hand, and the rational and striving will on the other. As Kane points out, I think we would all be happy with a purely deterministic relation from the rational to the striving will, but we are not all happy with a deterministic relation from the desiderative will to the rational will, for libertarians would take strong exception.

But note that our judgements of praise and approbation are sensitive too to this distinction between the desiderative will and the rational will. If a man is genetically disposed to a short temper, then we don’t blame him for that. We say something like, ‘He can’t help the desires he was born with [the desiderative will], but we can hold him accountable for what he does with those desires [the rational will] (so long as they don’t compel him)’. The distinction here concerns what an agent is given, what he finds himself with, and *what he does* with what he has been given. The desiderative will concerns the former, and the rational will the latter. This distinction is one we all intuitively acknowledge in our assessments of moral responsibility.

But once armed with this distinction I think we have all we need to state a plausible classical compatibilism. We say that f_m —the modal restriction found in PAP—is a function that moves from a world to a set of premises that offers a complete description of the relevant agent’s desiderative will prior to their choosing. The central suggestion of my paper can therefore be put like this: to claim that ‘S could have done otherwise’ in the sense required for moral responsibility is just to claim that S’s intending otherwise was compatible with their desiderative will. More formally, that

$$(5) \sim(\text{S's desiderative will}) \Rightarrow \sim(\text{S intends otherwise}).$$

⁹ It may be that the distinction between the rational and the striving will doesn’t amount to much. It may be only the distinction between deciding now to do something now and deciding now to do something later.

And to say that ‘S couldn’t have done otherwise’ means

(6) S’s desiderative will $\Rightarrow \sim$ (S intends otherwise).

I think we can see the force of this idea when we consider scenarios where we would be inclined to describe the desiderative will as entailing the rational. Consider a case of a man with a strong addiction to heroin; in fact, his addiction is so strong that it is literally irresistible. We don’t wish to hold a man responsible for what he does under irresistible compulsion, and my account explains why: in such a case it looks as if what the agent chooses follows inevitably from his desiderative will, his addictive desires. Intuitively speaking, these make choosing otherwise impossible relative to the desiderative will and so remove his moral responsibility.

So far so good, but in fact (5) and (6) are too weak. A description of the desiderative will by itself doesn’t really entail very much. We want it to be the case that the addict’s decision is entailed by his desiderative will. But it won’t be, for without any beliefs about how to fulfill one’s desires, even an overwhelming desire can find no outlet in action. So we should include a description of the agent’s doxastic states alongside the description of their desiderative will. For the sake of completeness, we might also want to include the agent’s experiential states as well, in case we think they might have some bearing. I will suppose that the doxastic, desiderative and experiential states of the agent constitute the agent’s *psychological setup*. I intend the psychological setup to include all those aspects of the agent’s makeup that are commonly supposed to be the ‘springs of action’—beliefs and desires, even if they are not enough to provide an analysis of decision, are often considered adequate to account for all those features explanatorily relevant to the agent’s decision-making in ordinary cases.¹⁰ So, we should therefore replace ‘S’s desiderative will’ in (5) and (6) with ‘S’s psychological setup’.

Is the psychological setup enough to get us the required entailment

¹⁰ I am not particularly concerned to limit the psychological setup of the agent to beliefs, desires and experiences. If it should emerge that it would be better for my account to include yet more aspects of the mind as part of the psychological setup, then I am happy for the psychological setup to be expanded as required—with one proviso: that it does not include what I introduce below as the ‘individual nature of the agent’.

from addictive desires to acting upon such desires? It is not, because it may be that the addict ceases to exist before any choice issues from his mental faculties. So, we need to include S's continued existence (E) alongside a description of S's psychological setup. But including S's continued existence in (5) and (6) isn't enough either. For suppose that an addict is going to act on an irresistible desire. We want it to be impossible for him to do otherwise. However, again, it won't be, because it may be that an external force (an angel swooping down from heaven, say) interferes and causes the addict to intend otherwise than that which the desiderative will would have determined. So, we must also include the claim that the agent is not interfered with in such a way. All I intend to include in this non-interference claim (NI) is the idea that no force external to the agent and the agent's psychological setup is directly causally responsible for the agent's producing his decision. The only admissible candidates are features intrinsic to the agent or his psychological setup. But continued existence (E) and non-interference (NI) aren't enough either; we must also include the laws of psychology (P), for without these there is nothing to rule out the possibility of an overwhelming desire to ϕ 's giving rise to ψ -ing rather than ϕ -ing.

The resulting analysis is therefore more complicated. Suppose S made a decision at t . Strictly speaking, f_m must generate all the above premises such that a more complete understanding of 'S could have done otherwise at t ' in the sense required for moral responsibility is this:

$$(7) \sim(\{S\text{'s psychological setup, E, NI, P}\} \Rightarrow \sim(S \text{ intends otherwise at } t)).$$

Spelled out:

An agent S could have done otherwise at t (in the sense required for moral responsibility) iff a description of S's desiderative will, and S's doxastic and experiential states, over an interval¹¹ immediately prior to t in conjunction with the claims that (i) S continues to exist until t and (ii) S's decision-making at t was not

¹¹ It does not matter which interval, for, even if the interval is very long, what S's psychological setup was like in the distant past won't be relevant to what it is possible for S to intend at t —what will be relevant is S's psychological setup as t approaches.

directly caused by any force external to the agent or his psychological setup, and with (iii) the actual laws of psychology, do not entail that S choose as he did at t .¹²

Is this the final analysis? I think we must make one more adjustment. (7) would give, I think, the wrong result in cases where the decision of the agent is the result of direct causal interference by an external force. If an angel swoops down from heaven and causes me to decide, in a manner that bypasses the psychological setup, to sing ‘God Save the Queen’, then wouldn’t I, if people complained about the noise and I were aware of what the angel had done, complain that I could not help it? But (7) would get the result that I could have helped it, because my not choosing to sing would be entirely consistent with my psychological setup, etc. just prior to the angel swooping down and causing me to decide to sing.

As a result, I think we must also suppose NI to be actually true, and the decision brought about in the proper manner for the ‘could have done otherwise’ claim to be true. So, in addition to the content generated by f_m , we will have to add the non-interference claim alongside it. The final analysis of ‘S could have done otherwise at t ’ is therefore:

(8) NI & $\sim(\{S\text{'s psychological setup, E, NI, P}\} \Rightarrow \sim(S \text{ intends otherwise at } t))$.

‘S couldn’t have done otherwise at t ’ is just the negation of (9):

(9) $\sim(\text{NI} \& \sim(\{S\text{'s psychological setup, E, NI, P}\} \Rightarrow \sim(S \text{ intends otherwise at } t)))$.

It might be thought that narrowing things down to whether or not an agent can intend otherwise given his psychological setup is too restrictive a model to handle the great variety of moral-responsibility contexts. Suppose a mugger holds a gun to my head and suggests it would be prudent for me to relinquish possession of my immediate finances. I agree and hand my wallet over. If, after the fact, someone

¹² For those desirous to know how my account would be incorporated into Kratzer’s later framework, it is achieved as follows: let the propositions detailing S’s instantiation of his psychological setup be the modal base. E, NI and P are to be considered as the ordering source.

were to challenge the propriety of my doing so, I might well respond that ‘I couldn’t have done otherwise’. Such an expression would be strictly false on my account: after all, I could have formed the intention to fight the mugger. Or suppose I am tied up and a child drowns in front of me as I watch helplessly. I would again complain that I couldn’t have done otherwise. But this will likewise be false on my account: I could have intended differently.

One might think, therefore, that an account of the modal restriction that varies more widely in generated content given context would be desirable. In the mugging case, one might say, the modal restriction wouldn’t generate a description of my desires, but instead a set of rules about what are reasonable courses of action given the value of one’s life. In that case, what I mean when I say ‘I couldn’t have done otherwise’ is that my resisting the mugger was not possible relative to those rules. In the case of the drowning child, it might be suggested that the modal restriction would generate a list of all possible bodily exertions in that situation, such that the ‘I couldn’t help it’ claim expresses the conviction that relative to my position and the laws of nature, there was no possible bodily exertion that would have resulting in the breaking of my bonds.

Such an account might, at the end of the day, be a better way to go. There would be no type of modal restriction distinctive to moral-responsibility contexts in that case.¹³ Nevertheless, my more uniform account can still handle these sorts of examples. In the mugging case, what I really mean when I say that ‘I couldn’t have done otherwise’ is that I couldn’t have *reasonably* done otherwise. In other words, yes, I could have done otherwise, but it wouldn’t have been reasonable to do so (because I would have been shot). In the drowning child case, when I say that ‘I couldn’t have done otherwise’ I am again speaking loosely. Indeed, strictly speaking, I could have done otherwise, but what I mean is that, of all the things I could have intended, none of them would have had the desired result (the rescuing of the child).

Whatever the advantages of variant accounts, it is the analysis given in (8) and (9) that I am defending on this particular occasion. It seems to get us what the compatibilist wants. It appears sufficient to

¹³ I thank Pablo Rychter for this suggestion.

ensure that the addict cannot do otherwise, yet doesn't require anything like the falsity of determinism. Furthermore, it avoids all the old objections to CAB. I have not analysed the 'could have done otherwise' claim as a subjunctive conditional, and therefore the objections to the old classical compatibilism aimed to exploit this feature cannot be deployed against my proposal. I have made the suggestion that the alternative possibilities requirement on moral responsibility is concerned with *compatibility* or *compossibility*, rather than the truth of various 'would' conditionals: broadly speaking, just so long as an agent's desiderative will is compatible with one deciding to do otherwise, then one 'could have done otherwise' in the sense required for moral responsibility. Such, I suggest, is the compossibility that PAP is fundamentally concerned with.

5 Frankfurt and other considerations

Why might compatibilists have overlooked this suggestion? I think there are two reasons. Firstly, compatibilism has often been understood, both by its proponents and its opponents, as precisely the claim that the rational will is fixed by the desiderative will! The traditional 'strongest desire wins' compatibilism of Hobbes (1651) and Edwards (1754), for instance, is a compatibilism that insists that our decisions *are* determined by what the understanding recognises as the strongest desire. I don't think that such compatibilisms have an adequate way of handling overwhelming desires—the addict, if literally overwhelmed by his desire for a drug, surely has his responsibility diminished, if not removed—so I don't consider it a problem that my way of securing PAP is inconsistent with their proposals.

Secondly, it might be thought that the phrasing I reach for is too close to libertarian phraseology for comfort. Locating the freedom of the agent in what the agent does with his desires and so forth looks like exactly the sort of thing libertarians have demanded throughout the years. Well, I would concede that the theory I give here is a concession of sorts to the libertarian: he was quite right to insist on a measure of independence for the rational will from the desiderative. It might then be wondered how I can concede this consistently with the truth of determinism. For if the desiderative will, or the psychological setup more broadly, doesn't determine the rational will,

then what does? But what lies between the psychological setup and the rational will? It is the agent. It is he who decides what is going to happen given the content of the desiderative will. Accordingly, I move that it is the individual nature of the agent that determines the rational will. I don't mean by 'nature' the essential properties of the agent—they might not be essential to either being an agent or being that particular agent—I merely intend to describe categorical properties of the agent distinct from the psychological setup that I think explain why one decision rather than another is issued in typical cases of decision-making.

This helps make it clear how my theory differs from the libertarian's; for let 'A' denote S's individual agential nature; then I am happy to agree that

$$(10) \{S\text{'s psychological setup, E, I, P, A}\} \Rightarrow \sim(S \text{intends otherwise at } t)$$

can be true consistently with the truth of 'S could have done otherwise' (in the moral-responsibility sense). But the libertarian denies this. For the libertarian insists that, broadly speaking, there must be no entailment from the psychological setup plus *anything* (or at least any hard fact¹⁴) to the rational will for the relevant possibility of doing otherwise to obtain. I contend for the weaker position that, speaking loosely, there must merely be no entailment from the psychological setup to the rational will. Nevertheless, still speaking loosely, I do believe that there is an entailment from the psychological setup *plus* the nature of the agent to the rational will.

Finally, let's deal with Frankfurt's celebrated counterexample to PAP. Frankfurt was trying to give us an example where an agent was morally responsible for what he did, but couldn't have done otherwise. I think it should be clear, relative to my account of 'could have done otherwise', that Frankfurt's scenario is one where Jones *could* have done otherwise. (8) is a conjunction, and so if Frankfurt's counterexample is to succeed, then at least one of these conjuncts will have to be false of Jones. Is the non-interference (NI) claim false? It is not. It is an important part of Frankfurt-style counterexamples that there is no actual interference by Black or his device at all. Such

¹⁴ See Plantinga 1986 for an explanation of the distinction between hard and soft facts.

things turn out to be entirely unnecessary in securing Jones's decision. Is it the second conjunct that is false? Is there any entailment from Jones's psychological setup, in conjunction with his continued existence, his not being interfered with, and the laws of psychology to his not doing otherwise than deciding to go ahead and kill Smith? We have no reason to think so. We would have, if Jones acted from an addiction or compulsion. But again, Jones's decision is supposed to be (aside from its grim objective) a regular and ordinary piece of decision-making. Jones is not supposed to be driven along by an overwhelming inferno of desire. We do not, therefore, have any reason to think that his decision to go ahead and kill Smith is entailed in the relevant way, quite the contrary.

Nor does Black's brooding presence over Jones's activity do anything to affect this lack of entailment. The only way in which we could get the conclusion that Jones couldn't do otherwise (in the relevant sense) on account of Black is if the set of premises f_m points us to includes a description of Black, his behaviour, his device, and his motives. But my account is of a fundamentally abstractive character: I am interested in whether or not the agent's psychological setup, *considered in the abstract*, is consistent with a different intention being formed. So, f_m will never stray so far outside the agent as to include someone else and their schemes in the set of premises it points us to, and for this reason Black will always be irrelevant, and all such Frankfurt-style counterexamples will fail to counter.¹⁵

Matthew J. Hart
 Department of Philosophy
 Mulberry Court, Mulberry Street
 University of Liverpool, Liverpool
 England
 L69 7ZY
 m.hart@liv.ac.uk

References

Austin, John Langshaw. 1961. Ifs and cans. In *Philosophical Papers*. Oxford:

¹⁵ I'd like to thank Simon Kittle and Daniel Hill. Their comments on an earlier draft saved me from various blunders.

- Oxford University Press.
- Bailey, Andrew. 2012. Incompatibilism and the past. *Philosophy and Phenomenological Research* 85(2): 351–376.
- Berofsky, Bernard. 2006. Classical compatibilism: not dead yet. In *Moral Responsibility and Alternative Possibilities*, ed. by David Widerker and Michael McKenna. Aldershot: Ashgate Publishing Limited.
- Dennett, Daniel. 1984. I could not have done otherwise—so what? *Journal of Philosophy* 81(10): 553–565.
- Edwards, Jonathan. 1754. *The Freedom of the Will*. Various editions.
- Fischer, John Martin and Ravizza, Mark. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Frankfurt, Harry G. 1969. Alternate possibilities and moral responsibility. *Journal of Philosophy* 66(3): 829–839.
- Hobbes, Thomas. 1651. *Leviathan*. Various editions.
- Kane, Robert. 1998. *The Significance of Free Will*. Oxford: Oxford University Press.
- Kenny, Anthony. 1975. *Will, Freedom and Power*. Oxford: Basil Blackwell.
- Kratzer, Angelika. 1977. What must and can must and can mean. *Linguistics and Philosophy* 1: 337–355.
- Kratzer, Angelika, 1981. The notional category of modality. In *Words, Worlds, and Contexts: New Approaches in Word Semantics*, ed. by Hans. J. Eikmeyer and Hannes Rieser. Berlin: de Gruyter.
- Kratzer, Angelika. 2012a. *Modals and Conditionals*. Oxford: Oxford University Press.
- Kratzer, Angelika. 2012b. What must and can must and can mean. In *Modals and Conditionals*. Oxford: Oxford University Press.
- Lehrer, Keith. 1968. Cans without ifs. *Analysis* 29(1): 29–32.
- Lewis, David. 1976. The paradoxes of time travel. *American Philosophical Quarterly* 13(2): 145–52.
- Maier, John. 2013. The agentive modalities. *Philosophy and Phenomenological Research* 90(1): 113–134.
- Plantinga, Alvin. 1986. On Ockham's way out. *Faith and Philosophy* 3(3): 235–269.
- Widerker, David and McKenna, Michael. 2006. Introduction. In *Moral Responsibility and Alternative Possibilities*, ed. David Widerker and Michael McKenna. Aldershot: Ashgate Publishing Limited.
- Van Inwagen, Peter. 1984. Dennett on 'could have done otherwise'. *Journal of Philosophy* 81(10): 565–567.
- Vetter, Barbara. 2015. *Potentiality*. Oxford: Oxford University Press.