

FREE WILL AND MODAL RESPONSIBILITY

WILLIAM BONDI KNOWLES

University of Warwick

In the last half-century increased awareness of modal issues has been brought to bear on the free will debate. It has been argued that the context dependence of possibility claims can be exploited to mount a defence of compatibilism, the idea being that the kind of possibility to do otherwise ruled out by determinism is distinct from the kind of possibility to do otherwise needed for free will. The potency of this idea, however, is still under-appreciated. It is often confused with conditional analyses of alternative possibilities, and many assume that the forms of possibility the compatibilist points to are somehow less “categorical” than the incompatibilist’s preferred *all-in* possibility. Moreover, Christian List’s questionable agent-level compatibilism has recently become the main representative of the idea. In fact what is needed — so it is argued here — is to combine increased modal awareness with the traditional compatibilist picture of the relevant freedom being freedom from external compulsion.

1. Introduction

Here is an informal statement of the consequence argument for incompatibilism. If determinism is true, then by definition it follows from a past state of the world and the laws of nature how any given future state will be, including how a given agent will act in that future state. It is not possible for the agent to break the laws of nature or change the past. Therefore the agent is unable to do otherwise than the laws and past determine him to do: the agent lacks alternative possibilities. Hence the agent has no free will. Hence incompatibilism.¹

It is now recognised on both sides of the debate that at most this argument shows that determinism is incompatible with a specific kind of possibility to do otherwise.² The incompatibilist has no grounds for denying that there are

1. The reasoning is traceable to van Inwagen’s “main argument” (1975:191) and finds support to this day, see e.g. Capes (2019).

2. Grzankowski (2014) is an example of someone both sympathetic to the consequence argument and aware of this point.

Contact: William Bondi Knowles <wbk1994@gmail.com>

possible worlds (so to speak) with either different laws or different pasts than our own, nor for denying that in some such worlds the agent does otherwise than here and in that sense has alternative possibilities. What the argument perhaps shows is that it is not possible for agents to do otherwise *given* the actual past and natural laws, the least tendentious name for which is “all-in” impossibility to do otherwise.³ All-in possibility, e.g. biological possibility, is a form of what is commonly referred to as *relative* or *constrained* modality. It is created by “holding fixed” certain actual facts when asking what is possible, and as relative/constrained possibilities go it is highly restrictive: one holds fixed *all* natural laws and *all* facts of a given past state—down to the very last detail.

This easily-missed aspect of the consequence argument raises the prospect of a compatibilist reply to it: deny that the kind of alternative possibilities needed are of the “all-in” variety. The reply has the benefit that it does not rely on denying that some form of alternative possibilities may be needed for free will, nor on claiming that we can change the past or laws of nature. It first saw a flurry of attention in the seventies⁴ and now again after recent work by Christian List, and the aim here is to lend it further support. To this end three argumentative sub-objectives will be pursued.

First, in §2, some reasons are considered for thinking that the relevant kind of possibility for free will is all-in possibility. Issue is taken with the common idea of all-in possibility as a privileged sort of “real” or “categorical” or “tout court” possibility. At best, there is some justification for regarding all-in possibility as a kind of default, to be ditched only if the compatibilist can provide a rationale for doing so.

Second, in §§3–4, the search for such a rationale—and with it a different, compatibilism-friendly form of possibility—begins. An assessment is given of List’s recent claim that the possibility relevant to free will holds fixed the higher-level, “agential” facts rather than the lower-level, microphysical facts. The claim is found lacking in argumentative support, but the things List does say point the way to a more promising alternative.

Finally in §§5–6 this alternative is outlined and argued for. The claim will be that in typical free will contexts it seems practical to not hold fixed everything about the situation, but rather to leave a gap where some of the agent’s internal psychological traits should be (such as desires, values, etc.). Consider moral responsibility. A lack of alternative possibilities is a good excuse for having acted a certain way partly because then one can say things like “Anyone would have done the same in those circumstances.” If the circumstances cited were the “all-in” circumstances, which include both the external circumstances

3. The name ‘all-in’ may originate with Austin 1956. It could also be called Leibnizian possibility after its early proponent in free will contexts (see Leibniz cited in Chisholm 1964:13).

4. E.g. Kratzer (1977), Horgan (1979), and Lehrer (1976).

and the internal traits of the agent, then this would amount to saying “Anyone *exactly like me* would have done the same in those external circumstances,” the exculpatory effect of which, reflection will show, is zero. For the excuse to make sense it needs to be impossible to do otherwise given the *external* circumstances specifically. The view arrived at thus combines traditional compatibilism (that the relevant freedom is freedom from external compulsion) with the more recent appreciation of the context dependence of modal claims.

I do not claim that the view arrived at is particularly original. The idea that one should abstract away from the internal characteristics of agents when asking what is possible for them is relatively frequently mentioned in the literature. But I do hope to build a convincing argumentative narrative for the view, situated in the contemporary debate. I also want to clear up some modal misunderstandings along the way, starting now. For although the view has been mentioned, care is not always taken to make the necessary modal distinctions. The view is often assumed to be committed to the conditional analysis of the ability/possibility to do otherwise, and thus to inherit its problems.⁵ That is a mistake. The view is that the possibility to do otherwise relevant for free will holds fixed only the external facts of the situation. The conditional analysis, on the other hand, is that the possibility to do otherwise consists in that one would have done otherwise if one had chosen to (or wanted to, or some such).

These are inequivalent, and importantly so. Having conditional freedom, i.e. an action being such that one would do the action if one tried to, does not entail that one is free from external compulsion: such conditionals can be true even for actions one’s circumstances prohibit. Suppose someone is being manipulated (by hypnosis or neurological implants or whatever) into not wanting to go for a walk. Then it can still be true that they would go for a walk if they wanted to. But they clearly aren’t free to go for a walk, because they’re not free to want it. This sort of case is commonly cited as a counterexample to conditional analyses of freedom.⁶ But it isn’t a counterexample to the view pursued here, because in this case one isn’t free from external compulsion. So not only is it wrong to equate the view pursued here with the conditional analysis, it also obscures an important advantage of the former vis-a-vis the latter.

2. Is All-in Possibility “Categorical”?

All-in possibility holds fixed all of a past state of the world and all the laws of nature. Does the incompatibilist have any positive reason for thinking that this

5. E.g. Bok (1999:94). Also see Watson (2004) and the varying descriptions on e.g. pages 91 and 175.

6. See e.g. Chisholm (1964: 6), Horgan (1979: 345). Whittle (2010) argues that similar problems affect more sophisticated versions of conditional and dispositional analyses of “could.”

should be the relevant kind of possibility for free will? It has been argued that at this stage of the debate, the moral question gains primacy and that in order to work out which kind of possibility is relevant to free will, one first needs to consider which kind of possibility will be relevant to moral responsibility.⁷ Ultimately moral considerations will become salient, but I set them aside for the time being. There are available moves which don't involve morality, and these also need to be considered.

There are three broad ways for arguing in favour of all-in possibility. The first two are unsuccessful. The third, however, is both sound and valuable for the debate since it lends incompatibilism some initial credibility. But its scope is limited: it may establish all-in possibility as a kind of default for free will contexts; but it does not call off the search for alternatives.

The first argument is that we should hold fixed the past plus the laws of nature because we mere humans are incapable of changing them. Precisely why this incapability should justify all-in possibility has not been stated explicitly, but it seems to be what many incompatibilists may have had in mind.⁸ Perhaps the idea is that it isn't *really* possible for an agent to do otherwise if the agent only does otherwise in possible worlds with different pasts or laws of nature, because then in order to do otherwise the agent would need to change them, and that the agent cannot do. Possible worlds with different laws/pasts therefore do not represent genuine possibilities, and we should hold fixed the actual laws and past.

But this argument is unsound. It is false that in order to do otherwise the agent needs to *change* the past or laws of nature. There are possible worlds where the agent does otherwise and the past/laws just *are* different. It would only be true that the agent would need to change the laws of nature if the agent's actual past/laws were held fixed, which is precisely the question at stake.

It can't be denied that the impossibility of changing the past/laws has some relevance to what one should hold fixed when considering an agent's possibilities. If there were no such impossibility, if we could edit the past/laws like Word documents, then there would be little sense in holding them fixed. Any resulting impossibility would be of a trivial sort since one could overturn the conditions being held fixed; it would be no more interesting than, say, the impossibility of your staying home tomorrow given that you will leave the house. So the inability to change the past/laws may well be a necessary condition on the reasonableness of holding them fixed. But it has yet to be shown to be a sufficient condition. It is hard to think why it would be.

Moving on, the second argument is that the more that is held fixed from the actual world, the more real the possibility becomes. All-in possibility is therefore

7. Franklin (2015: §5).

8. See e.g. Ismael (2016: 190), who calls it the "old" reason for keeping the past/laws of nature fixed.

“outright” or “categorical”⁹ possibility: it tracks what is *really* possible, and is therefore the appropriate kind of possibility for free will. My general objection to this is that if you keep holding more and more actual facts fixed, then the only sense in which the possible/impossible distinction becomes more real is that it more closely resembles the actual/non-actual distinction. But any interesting modality doesn’t just concern what is actual: it concerns the way the world could have been and could be, not just the way it was and will be. Consider the following way the incompatibilist’s position has been described:

Incompatibilists are often puzzled by the compatibilist’s focus on non-actual conditions. They are typically driven by doubts about whether it is fair to blame an agent if that agent could not have done otherwise given the way things actually were. The incompatibilist understandably struggles to see why we would spend our time focusing on non-actual scenarios, when these differ in crucial respects from the situation that actually occurred. (Elzein & Pernu 2017: 225)

I do not agree that the incompatibilist’s struggle is understandable. When the question concerns what is possible, it is hardly a valid complaint that one focuses on non-actual scenarios. The idea seems to be that if an agent does otherwise in some possible world than he does here, and if that world differs in some “crucial respect”, then really all that’s been established is a conditional possibility: that the agent could have done otherwise if those crucial respects had differed, and that if you want to know what the agent could *actually* have done, then you need to ignore such worlds. But that is just not true, at least not unless you’re already assuming that what is at stake is all-in possibility. Possible worlds represent actual possibilities; they represent what is possible *here*. It is a complete mistake to avoid considering alternative possible worlds when the question is what is possible.

The idea that all-in possibility is somehow “real” finds little support in the modality literature. Modal claims are usually divided into the absolute and the relative. How to make this distinction precise is disputed,¹⁰ but the standard view is that the absolute modal claims are those which hold nothing fixed; which take in the full complement of possible worlds. Whether a modal claim is absolute or relative thus concerns whether it is constrained by actual facts. It is absolute/unconstrained modality which is standardly regarded as the most real, since it concerns the kind of modality which doesn’t involve the distinctly human element of holding things fixed. This perspective lends no support to the ultimate reality of all-in possibility, which is a form of relative/constrained modality.

9. Respectively, Elzein & Pernu (2017: 240) and O’Connor & Franklin (2022: §2.2).

10. Some options are surveyed by Hale (2012).

One idea mentioned in the modality literature might sound conducive to seeing all-in possibility as “real.” A candidate for making sense of absolute necessities is that they remain necessary no matter what is held fixed.¹¹ Likewise, one might think (though I have not seen anyone say this) that an absolute possibility should be possible no matter what is held fixed. But this is hopeless: if one wanted to, one could hold all facts fixed, past, present and future. In possible worlds terms, one could restrict attention to a single possible world: the actual one. This definition would therefore make “absolute possibility” coextensive with the actual truth, independently of the truth of any determinism thesis. (Maybe someone would object to this point on the basis that they reject facts about the future. But if that rejection is motivated by a worry that future facts are incompatible with free will, perhaps made vivid by invoking a fore-knowing deity, present considerations point to a less dramatic solution: deny that free will’s possibility is all-in.)

In sum, neither of the first two reasons succeed. But there is a third which, though limited in ambition, fares better. The kind of possibility relevant to free will is *situated*, in the sense that it concerns what is possible given the agent’s situation and circumstance, rather than, say, nomological possibility, which only holds fixed the laws of nature and seems far too broad: it may be both nomologically possible for me to go the shop tomorrow and nomologically possible for me to do otherwise, but this doesn’t say very much given that it is also nomologically possible for me to, say, run a hundred meters in 9.5 seconds, or give an off the cuff two-hour lecture on Thomas Aquinas’s theory of logic, or what have you. What nomological possibility lacks is facts about the agent’s situation. Another way of seeing that the relevant form of possibility is situated, is by considering an apparently legitimate way of formulating the possibility involved in my free choice regarding whether to stay home or go to the shop. On this formulation, both possibilities are open to me in the sense that both are possible continuations of the actual history of the world: I am at a “branching point” between staying home and going to the shop. Again, this would make no sense without taking into account what the actual history is. The relevant form of possibility is therefore situated.

The fact that the relevant possibility is situated doesn’t prove incompatibilism: there is a difference between holding *some* facts of the situation fixed and holding fixed *all* of them. But it does enable the incompatibilist to say that any kind of possibility holding fixed only an incomplete description of the situation—either because it is generally undetailed or misses out specific aspects—requires motivation. Until or unless such motivation is provided, the default position is to take the whole situation into account.

11. Hale (2012: §1).

The burden of proof is levered onto compatibilists. Attention now turns to whether they can shoulder it and, first, to what conditions need to be met by the account offered. The main victory of this section, however, is compatibilist. No support has been found for the idea of all-in possibility as holding some privileged “real” status, deviation from which will only ever be a thin and superficial theoretical construct. Given the frequency with which ‘all-in’ is described as ‘categorical’ or ‘outright’ in the literature, a significant part of the lingering appeal of incompatibilism appears undermined.

3. Conditions on a Compatibilist Account

Compatibilists need to motivate a less restrictive form of possibility than all-in, but not any motivation will do. It also needs to be of the right kind. Consider one possible motivation which might initially seem sufficient but which on further reflection turns out, I think, not to be. I have in mind the claim that we don’t typically employ all-in possibility in everyday circumstances, not even in typical blame/praise contexts, e.g. Kratzer’s claim that a judge asking himself whether a murderer could have acted otherwise doesn’t mean to hold fixed the whole situation of the crime, but rather certain aspects of it.¹²

Even if the reader finds the claim plausible, it isn’t automatically the compatibilist victory it might seem. One thing is to deny that we typically employ all-in possibility in our everyday lives, another is to use this as an argument for free will. That is a further step, and whether it is appropriate depends completely on *why* we don’t employ it. It might be thought—though I am not in any way endorsing this idea, its inclusion is purely for illustrative purposes— that we don’t hold fixed everything about a situation because of the fairly obvious epistemic problem. A complete description of a given state of affairs encompasses an absolutely enormous amount of information, far more than any living human could get their head around, most of it completely trivial. And at the micro-level of electrons and atoms our knowledge scarcely extends to what kinds of things there are, let alone to descriptions of individual particles’ movements in our environments, still less to predictions of future movements. It might seem purely a matter of epistemic convenience, therefore, to hold fixed a rough picture of the situation rather than the all-in description.

But if this was the only reason for shunning all-in possibility, that would be insufficient for a robust notion of free will. Surely genuine free will is not the result of some epistemic limitation, as if we might not even bother with it if only we knew more. Freedom is not contingent on ignorance. Nor is moral

12. Kratzer (1977: 343).

responsibility. If this were the only kind of free will compatibilism could establish, compatibilism would be doomed.¹³

It is therefore insufficient to claim *that* we don't employ all-in possibility in everyday modal circumstances. It also matters *why* we don't. It might be useful to sum up the lessons from this section and the last in some conditions the compatibilist argument must meet to be successful.

- (a) A rationale is needed for avoiding all-in possibility in favour of a less demanding alternative.
- (b) The rationale cannot be epistemic or otherwise insufficiently robust.
- (c) The less demanding alternative possibility needs to be a form of *situated* possibility and consider the facts of the agent's specific situation.
- (d) The alternative needs to be compatibilist, i.e. at least sometimes we need alternative possibilities even if determinism is true.

The prospects for meeting these conditions will now be explored, beginning with Christian List and his agent-level possibility.

4. Agent-level Possibility

List begins by stipulating that when asking what is possible at the agent-level, one holds fixed all the higher-level facts (about beliefs, sticks, stones, etc.) but none of the lower-level facts (about neurons, atoms, etc.). He has been interpreted, wrongly, as claiming that at the agential level one holds fixed *only* the psychological facts of the agent.¹⁴ This is clearly not what he intends given that he describes the relevant set of possible worlds as follows:

Let *S* be the set of all possible agential states, that is, the set of all possible states of the relevant agents *and their macroscopic environment* as specified by our best higher-level theory of human agency. (2014: 164, my emphasis.)

With this definition of the agential level, List claims first that indeterminism at the agential level is compatible with determinism at the lower level, and second that agent-level possibility to do otherwise is what is needed for free will. The first claim is not implausible and will be granted here, even though its depen-

13. Elzein & Pernu (2017) use this point to object to List's compatibilism.

14. See Kaiserman (2022: 1315), who is, incidentally, perfectly aware of the definition and yet chooses to relegate it to what List says "at some points" rather than simply being his view. Given that holding fixed only the psychological facts would be an easily refutable position, as Kaiserman is quick to point out, this interpretation is uncharitable.

dence on non-obvious metaphysical theses (e.g. the multiple realisability of higher-level facts by lower-level facts) seems somewhat in tension with the spirit of compatibilism. It is the second claim that's most interesting—and problematic. Much of what follows is quite critical of List, but to avoid misunderstanding it should be kept in mind that this is with a view to improving upon the general kind of view he defends, not refuting it. I agree with him about whether to be a compatibilist and broadly how to be.

List's argument is that we should accept agent-level possibilities because of their use in behavioural science:

A central concept of any version of decision or game theory ... is an agent's set of possible actions or strategies. ... Sometimes the addition or removal of options can make a significant difference to what the agents are predicted to do even if these options are not ultimately chosen. Unless we accept that there is at least a thin, technical sense in which such options could have been chosen, it is hard to make sense of those effects. (2014: 170)

If agent-level alternative possibilities really are indispensable to folk/scientific practices of predicting and explaining human behaviour, then some rationale is gained for adopting the less restrictive agent-level possibility rather than all-in. And it seems right that behavioural theorising involves alternative possibilities *of some kind*, both when constructing general theses and in everyday behavioural predictions. If we want to predict what flavour ice cream an imagined agent Freddy will buy from the van, for example, it is crucial information what Freddy's options are. This is also a specifically agential prediction: when weather forecasters forecast the weather, they don't take into account the options the weather has. Presumably this reflects that agents do in fact consider their possible options and deliberate which one to choose, and thus this links to an intuitive aspect of free will: that in choice contexts we see possible alternative courses of action as branching out ahead of us.

There is, however, a highly significant hole in List's argument. There seems to be no connection between the theoretical usefulness of several available possible courses of action and specifically *agent-level* possibilities. Such possibilities hold fixed *only* macro-level facts and *all* macro-level facts. But the usefulness of alternative possibilities in behavioural theorising justifies neither the *only* nor the *all*.

The *only*. The assumption behind keeping fixed only agent-level facts seems to be that behavioural sciences aren't concerned with the microlevel of atoms and neurons: they would work just as well if we had no such knowledge at all, therefore we should ignore lower-level facts when determining an agent's options. But

this reasoning would prove too much. The behavioural sciences aren't concerned with the existence of cars either, yet the existence of cars is held fixed at the agential level. Admittedly, as List has pointed out, macrolevel facts like the existence of cars are sometimes assumed in behavioural theorising when addressing certain choice contexts, e.g. when the person is deciding which mode of transport to use.¹⁵ But this will not provide the motivation he needs for the different treatment of facts from different levels, because it is equally easy to think of cases where behavioural theorising would take into account microlevel facts. Suppose someone is choosing between which of two pieces of shiny metal to steal, and that the piece to the left has the micro-properties of gold and the other not, and that the person knows this. When predicting or explaining the person's choice, this microlevel information would obviously be just as appropriate to take into account as the existence of cars for predicting a choice of a mode of transport. The ban on holding fixed microlevel facts in free will contexts has therefore not been sufficiently motivated, and in fact, a variation on the gold example seems to show that the ban is positively inappropriate. Consider the two pieces of shiny metal again, but this time instead of asking what is relevant for predicting what the person will actually do, the question is what the person *can* do; what his options are. Which piece of metal the person is free to bend with his teeth? Again, there is nothing inappropriate about taking the microlevel properties into account when answering this. And this shows that the micro-level facts often are relevant, contrary to List's claim, to determining what an agent's options are.¹⁶

Perhaps List would concede that some microlevel facts can be held fixed and yet resist that all of them should be, but then a principled reason would be needed for separating which microlevel facts to hold fixed and which to vary, the provision of which would in effect amount to a new positive story. Without one, the obvious conclusion would be that too much was left unsaid.

The *all*. When it comes to holding fixed all the macro-level facts, the rationale is even thinner. This is less problematic for List in his debate with incompatibilists, who will grant that one should hold fixed all the macro-level facts (since they think one should hold fixed all facts of any kind). But it is interesting either way—particularly with one eye on giving an alternative positive story—that the theoretical practices cited in support of agent-level possibility actually appear to point away from holding all the macro-level facts fixed. Take the case of prediction and suppose we want to predict which ice cream Freddy will buy from the van (similar remarks apply, *mutatis mutandis*, to explanation). The first step of this prediction is to establish Freddy's range of possible choices, his options. Which information is relevant for determining the options? If List were right,

15. List (2019: 91–92).

16. Similar objections are given by Elzein & Pernu (2017: 234), Weir (2016: 57), and Wright (2022).

then the answer would be: everything, or at least everything at the macrolevel, which might here be taken as the same thing. This includes which flavours the van offers, whether there is another van nearby, how much money Freddy has, Freddy's tastes and preferences, and so on.

I do not deny that Freddy still has remaining options after all this information is taken into account—doing so would amount to affirming macrolevel determinism, which I take no stand on. So suppose Freddy does have alternative agent-level options, as List thinks. At the next stage of the predictive process, one would expect these options to be weighed up against each other with the aim of reaching a prediction. But given that all the information has been used to determine the options, on what basis can we make the comparison? Indeed, if there are any options left after all the facts are taken into account, then that seems like a distinctly *bad* thing for the prediction. Once all the information is taken into account, there should ideally be only one remaining option, which can then be selected as the predicted purchase. That is, agent-level alternative possibilities may exist, but they don't seem to be the sorts of possibilities which are good for the predictive process.

List would, I think,¹⁷ respond by denying that it needs to be a bad thing for a prediction that multiple options remain once all the information is taken into account. He would say that in such circumstances a single option can still be pinpointed as being *rational* for the agent to choose. In other words, the agential facts don't entail which option will be chosen, but they can entail which option is rational. This point can be accepted, and it shows that prediction is compatible with the agent having several remaining options even when all the information is taken into account. But as List himself makes plausible, the positing of alternative courses of action isn't just compatible with the predictive process, it is also useful and maybe even indispensable for it. Clearly, however, it would be even better if there was only one remaining option when all the information was taken into account. This would make for a safer prediction.

To press the objection, consider that a consequence of List's view is that it would be a devastating blow for behavioural sciences if it turned out that macro-level determinism was true, because then one would not have alternative agent-level possibilities, and List sees such possibilities as essential to the behaviour theoretical process. This consequence seems implausible: why should behavioural theorists be worried about the existence of deterministic behavioural laws? If such determinism were established, behavioural theorists would presumably not lay down their tools but try to work out what those laws were. The apparent implausibility of this consequence means List needs some good positive reason for thinking that it is specifically agent-level possibilities which are relevant to behavioural theory.

17. This is his reply to a similar objection, see List (2014: §8).

I do not see that he provides any. Instead his arguments seem to establish only that *some* kind of alternative possibilities is needed. He claims, plausibly, that a lack of alternative options would trivialise certain non-trivial theses in decision theory. For example, if an agent only had one option in any given circumstance, then it would follow logically that a principle such as “An agent will always choose the option with the most expected utility” would be true. And the following absurd principle would be trivially true as well: “An agent will always choose the option with the least expected utility.”

This is a strong argument for the usefulness of alternative possibilities in decision theory, but not for specifically agent-level alternative possibilities. As List himself concludes, there must be alternative possibilities “in some sense of ‘possibility’.”¹⁸ Since List provides no convincing *pros* and there seem to be serious *contras*, it is reasonable to conclude that the possibility involved in behavioural theorising is not constrained by all the macro-level facts.

If List’s agent-level modality is dropped, what might an alternative be? Consider again the case of predicting which ice cream Freddy will buy. The predictive process begins by establishing Freddy’s options, and a sensible first pass might be that they are simply the flavours the van offers, which is approximately what Freddy himself will regard as his options. Once these initial options are established, the second step is to use the facts about Freddy’s personal circumstances—his preferences, parental instructions, boldness, etc.—to compare the possible flavours against each other, thus whittling down the initial set of possibilities to as few as possible, ideally just one (assuming he’s only after one), which will then be the prediction. This two-step process is what seems to be going on in decision and game theory: the desirability of each alternative isn’t typically regarded as determining which possible options an agent has, but rather as determining which of the options the agent will choose.¹⁹ For example, a classic problem of decision theory is what happens if the agent has at least three options and circular preferences, i.e. they prefer A to B, B to C, and C to A. Here the options—A, B and C—are set out *before* the agent’s preferences are taken into account.

In sum, List’s claim that alternative possibilities are crucial to behavioural theorising may be true, but it seems implausible that the form of possibility involved should hold fixed all the macro-level facts. That is obviously not good news for the specific agent-level view. But for the general project of reconciling the possibility to do otherwise with determinism, the prospects seem brighter. Since the theoretical practices don’t use all the information to determine the options, that equally suggests that the relevant kind of possibility is not of the all-in variety.

18. List (2019: 99, 101).

19. In fact, this is similar to how List himself describes the process of an “intentional explanation” (2019: 99).

5. The Compulsion Argument and “Gap-agential” Possibility

When considering which options were available to Freddy in the previous section, it seemed reasonable to hold fixed typically “external” circumstances like what flavours the van offered but not typically “internal” conditions like Freddy’s tastes and preferences. This suggests that a comeback might be on the cards for that most traditional of compatibilist ideas: the compulsion argument, according to which it is only when the facts are external to the agent that the agent’s free will is compromised. Thus one hears:

Freedom means the opposite of compulsion; a man is free if he does not act under compulsion, and he is compelled or unfree when he is hindered from without. (Schlick 1939: 7.6)

Where “from without” should be understood not merely as from outside the person, but as anything outside the person’s *will* specifically. A person’s will may be taken to comprise general properties like character traits and values and aims, as well as situation-specific desires and intentions. Facts about the person’s physique (such as their capacity for running a hundred meters in under ten seconds) and non-volitional psychological facts (such as their speed at solving maths problems) thus count as external. The idea is that what is at stake is whether the will is free, hence a course of action will be an open option only when there is nothing from outside the will impeding it from taking that course of action. There is more to say about the issue of precisely where to draw the distinction (e.g. what to say about cases of “internal compulsion”); it will be returned to in the next section.

The importance of the compulsion argument is that—as is perhaps known but certainly underappreciated—freedom from external compulsion can be given an exact statement in terms of possibility.²⁰ Specifically: freedom from external compulsion is logically equivalent to what one might call the *gap-agential* possibility to do otherwise. When determining gap-agential possibility, one takes everything about the agent’s situation except the agent’s internal properties, where one leaves a gap or blank. Then one asks: given all of that external circumstance, is it possible for the agent do otherwise? If yes, then the agent is free from external compulsion in the sense that the action isn’t necessitated by the external conditions. If no, then not.

20. Schlick: “The absence of the external power expresses itself in the well-known feeling (usually considered characteristic of the consciousness of freedom) that one could also have acted otherwise. ... [T]he feeling never says that ... under exactly the same inner and outer conditions I could also have willed something else.” (1939: 7.6, emphasis mine). This is nearly right but not quite: the possibility to do otherwise is not just a felt possibility, it’s the real thing.

Given this equivalence, the rationale needed for gap-agential possibility, which is a less restrictive form of possibility than the incompatibilist's *all-in*, can be appropriated from the reasoning behind the compulsion argument. The aim now is to show how that can be done. There are two main arguments behind the compulsion argument: the general case, and the moral case.

The general case. A fundamental aspect of having free will is sometimes being in situations where one has to select a course of action from several available alternatives. According to the traditional compulsion argument, what constrains the space of options in such contexts is only the external facts of the situation, not the internal ones. The considerations of the previous section suggest that this is right. As was seen, the options posited for Freddy's ice cream purchase are naturally regarded as determined by the external facts like which flavours the van sells. Internal facts about Freddy's tastes and preferences are then used to weigh up the options (either by Freddy when deliberating, or by a third party when predicting). The intuition can be sloganised: *one's preferences help determine which option is chosen, not which options one has to begin with.* When pondering which of an array of options to select, we judge them on their desirability, comparing them with each other and ranking them. That they are options is determined *before* this judgement of their relative desirability occurs. The role of things like one's preferences and tastes and likes and dislikes is therefore not to set out the options in the first place, but to aid in the selection between them.

Some rationale for gap-agential possibility is thus its use in prediction and deliberation. When we think of free choice as dependent on being at a "branching point" in history, where one can choose which future timeline to follow from a range of alternatives, those alternatives are gap-agential. Is gap-agential possibility to do otherwise feasible as a definition of free will? It seems plausible as a necessary condition on at least the mentioned "branching point" conception of free will, according to which an action could never really be freely chosen if there were *no* other possible alternatives, regardless of whether it felt free to the person, and regardless of whether it was in accordance with the person's desires. (Whether this form of free will is needed for moral responsibility is of course a completely different question, see the discussion of Frankfurt below.) It is harder to judge whether having the gap-agential possibility to do otherwise is also sufficient for freedom. If determinism is false, then some events might be random and thus not externally determined, while still being such that an agent lacks the control over them needed for genuine freedom in any sense of the word. Suppose some outer-space chemical reaction is indeterministic—it is fifty-fifty whether it goes one way or another. Then, for any given Earth dweller, it is gap-agentially possible for them to act such that it goes one way, and gap-agentially possible for them to act such that it goes the other. Yet one hardly has a free

choice, nor any kind of choice, between these two alternatives. For this reason, Schlick's above claim that "freedom means the opposite of compulsion" might be somewhat quick. But even if gap-agential alternative options do not provide a definition of free will, for the purposes of the compatibilist argument the weaker conclusion suffices: that the kind of alternative possibility necessary for free will is gap-agential, not all-in.

The moral case. The gap-agential perspective will seem suspicious to some and uninteresting to others if it cannot substantiate the link between free will and moral responsibility (understood as moral blame- or praiseworthiness). It can, but due to Frankfurt's influential arguments care is needed in stating the case. In Frankfurt-style examples, there is an overdetermination in place such that the agent's internal characteristics suffice for the action at the same time as the external circumstances rule out doing otherwise, and the intuition is that the agent may still be blameworthy in such circumstances even if they couldn't (gap-agentially) have acted otherwise. I do not dispute the conclusion. But gap-agential alternatives can be *relevant* to moral responsibility even if they are not strictly necessary; relevant, that is, in a way the colour of the agent's socks is not. In particular, it seems that citing a lack of options might still be a valid excuse of one's actions. This, according to Frankfurt himself, is because when one cites it as an excuse, it is implicit that the lack of options is what explains one's having acted as one did. It is to say that one acted that way *only because* one lacked alternative options. Frankfurt thus concedes that alternative possibilities and moral responsibility may be connected by the following principle:

A person is not morally responsible for what he has done if he did it only because he could not have done otherwise. (1969: 838–839)

For Frankfurt, this principle is acceptable because if one does something only out of a lack of alternatives, then it is not the case that one does so because of, in effect, one's internal circumstances ('what he really wanted to do'). Even the "actual-sequence" theorists thus allow some role for alternative possibilities in their doctrines. And this role is all that's needed for a rationale for the gap-agential perspective to be materialise. Because for this principle to be true, I shall now argue, the relevant sense of 'could' needs to be gap-agential, not all in.

External compulsion compromises moral responsibility, according to the traditional compulsion argument's reasoning, because the compelled action fails to reflect anything about the agent's "individual manner of decision making."²¹ Observe thus the kind of excuse one is able to give if one lacks the gap-agential possibility to do otherwise:

21. Schlick (1939: 7.6).

I did as I did only because I couldn't have done otherwise given my external circumstances. Anyone else in the same situation, no matter how virtuous, would have done exactly the same. Therefore, nothing about me as an individual or my moral worth can be inferred from my doing what I did.

This excuse seems to capture the intuitive exculpatory effect of lacking alternative options. When no judgement on the agent's character or moral worth is warranted by the deed, then it is hard to see how the agent can be justifiably blamed. It would not be reasonable to consider the action as a reason to hold a negative opinion of the agent.²²

Now compare with the excuse one might gain from lacking all-in alternative possibilities:

I did as I did only because I couldn't have done otherwise given my external and internal circumstances.

In this case, the person's internal factors may well have been instrumental in bringing about the action, and so a judgement on the person's moral worth may well be appropriate. Likewise, when 'situation' refers to the 'all-in' situation and thus includes *all* the facts both about the agent and the external setting, saying "Anyone else in the same situation would have done exactly the same" is obviously of little exculpatory value. It amounts to saying no more than: "Anyone exactly like me in exactly that external circumstance would have done the same." This reads more like a *mea culpa* than an excuse. It offers no reason at all not to let the action count as a reason to hold a negative opinion of the agent.

Summa summarum: gap-agential possibility is what is needed for alternative options in choice contexts. All-in possibility is not. Gap-agential possibility makes sense of the exculpatory effect of lacking alternative possibilities. All-in possibility does not. A rationale for avoiding all-in possibility and embracing gap-agential possibility is thus provided, and condition a) from the end of §3 is satisfied. The rationale isn't epistemic or otherwise inappropriately contingent, so b) is also satisfied. Condition c) is satisfied by it being situational: all of the facts of the agent's external situation are taken into account, right down to the micro-level. Finally, condition d) is satisfied because if only the external circumstances are held fixed, then the agent will often have the possibility to

22. "Source" incompatibilists of the kind swayed by typical manipulation arguments might regard this as a bad test for moral responsibility. They might say: "yes, this person's action suggests that they are wicked (or more neutrally, that they are disposed to do wicked things), but we can't blame them even so (if determinism is true)." Such statements seem to me contradictory. Either way, this is surely a good test of some kind of moral responsibility.

do otherwise even if determinism is true. This will be the case in exactly those circumstances where the agent is free from external compulsion.

6. The Internal/External Distinction

The view arrived at is close to the traditional compatibilist compulsion argument, the main upgrade being that freedom from external constraint is reflected in the perfectly real (albeit relative) gap-agential possibility to do otherwise. This simple trick solves one of the compulsion argument's main problems: how to make sense of the possibility to do otherwise without appealing to dubious conditional or dispositional analyses. Moreover, there seems to be a positive rationale for embracing gap-agential possibility and avoiding all-in possibility. The argument thus not only opens up a compatibilist path, it ushers us down it.

However, more needs saying about the internal/external distinction. Two problems in particular are worth addressing.

First, suppose determinism is true, i.e., that a complete description of a given state (including the laws of nature) entails the description of any future state. On this supposition, facts from the distant past—which by any metric are external to the agent—will entail the agent's current internal facts. Therefore, if determinism is true, then it is not possible to hold fixed all the external facts without automatically re-instating the internal facts. So the proposed "compatibilism," an opponent might point out, actually depends on the falsity of determinism, and is therefore not compatibilist at all.

This problem has an easy solution. I have argued that there is a good rationale for not holding fixed the agent's internal facts, relating to their desires, values, aims and so forth. This rationale is not affected by whether determinism is true or not. Hence, even if determinism is true, we should still not hold fixed those internal facts in free will modal contexts, which means that, if determinism is true, we should not hold fixed a complete description of the external facts from an ancient past state. All the external facts of a single state should only be held fixed once those can be contrasted with the agent's internal facts. So the solution involves a certain restriction on which historical external facts to hold fixed, but that restriction is not ad hoc or unprincipled: it has the same rationale as not holding fixed the internal facts in the first place.

It might be asked which ancient external facts should, if determinism is true, not be held fixed. I don't think it matters very much as long as it is enough to break the entailment to the current internal facts—one might hold fixed everything except a few key details or nothing at all. What is motivated by the arguments I have given is that all the external facts are held fixed once the contrast can be made with internal facts, and in particular at the time of the action.

Second, it has been objected that the internal/external distinction can't serve the compatibilist's purposes because of cases of apparently internal compulsion. At least some psychological traits, possibly including e.g. the addict's desire for drugs, the OCD sufferer's to wash their hands, the agoraphobic's to stay out of open spaces, do seem, intuitively, to constrain the agent's freedom. So it isn't the case that only external circumstances are held fixed in the modal context relevant to free will.²³

I think the compatibilist can accept that such psychological traits sometimes constrain freedom and deny that they are genuinely internal. In the moral context, a lack of alternative options is relevant, I have argued, because it means that the resulting action does not reflect any particular vice or moral defect in the person's character—even a saint would have done the same. Likewise, when cases of compulsive behaviours restrict freedom, the impulses do not reflect anything about the persons themselves (their "true" selves, as it were) or their normal manner of decision-making, but rather about the presence of an external affliction which is hi-jacking the person's control centre. This response is consistent with the original compulsion argument treatment of such cases. Regarding the free will of a mentally ill person, Schlick says that "we do not consider him free with respect to those acts in which the disease expresses itself, because we view the illness as a disturbing factor which hinders the normal functioning of his natural tendencies. We make not him but his disease responsible."²⁴

If this compatibilist diagnosis is correct, then behavioural disorders will not constrain the agent's freedom in those cases, which may well exist, where the disorders are part of who the person really is. I take this to be an acceptable consequence. It is only when the disorders disable the person's true nature that freedom is genuinely suppressed and moral responsibility diminished. A kleptomaniac who defended their stealing by saying "That's just the way I am" should not, on the basis of this defence, be excused.

Is the response adequate? The incompatibilist might reply that it threatens to collapse the internal/external distinction, since *all* of a person's internal traits ultimately have their causal root in circumstances external to that person, say from before they were born (at least if determinism is true). But this collapse does not occur. The point is that compulsive desires are external when they result from some external masking of the agent's true nature and behavioural dispositions. That is why the person's free will is suppressed. In cases where one's desires are causally traceable to factors before one was born, those factors can hardly be said to interfere with the agent's true nature, because at that time, there was no such nature with which to interfere.

23. Russell (1988: fn 8) calls this one of the most serious problems for the compulsion argument.

24. Schlick (1939: 7.6).

Alternatively, the incompatibilist might reply with a demand for an account of precisely when a desire is internal (i.e. belongs to the agent's true self) and when it is external. Giving such an account is of course no easy task; the most famous attempt to provide it is Frankfurt's appeal to second order-volitions, with its various offshoots,²⁵ and I have nothing better to suggest here. However, it seems to me that rejecting compatibilism for not having such an account is inappropriate, because everyone will agree, I think, that a person's desires can fail to properly reflect that person's true self under certain conditions. Consider just the case when one's desires are induced by hypnosis, or brainwashing, or, if possible, a neuroscientist's remote-controlled silicon brain implant. Everyone should agree that such desires are not representative of the agent's true self; therefore, the problem of providing an account which distinguishes internal from external desires belongs to everyone. The compatibilist's only special demand is that this account must be compatible with sometimes regarding "compulsive" impulses as external. I have no argument for thinking that this demand will be met beyond what I have already said, but then I also see no reason to doubt that it should be met. I therefore don't see the lack of an account of a person's true self as a reason to reject this form of compatibilism.

7. Conclusion

Some might find it hard to shake off the feeling that gap-agential possibility is somehow merely a "technical" or "thin" or "superficial" possibility, and that the foregoing therefore fails to engage properly with the consequence argument. In fact, gap-agential possibility is as full-blooded as relative possibilities can ever be: There is a clear rationale for what it holds fixed and what it doesn't. It is relative possibilities lacking such rationale, possibly including all-in possibility, that can truly be said to be thin and superficial. Whether or not determinism is true thus proves unrelated to the modal facts relevant to free will.

References

- Austin, J. L. (1956). Ifs and cans. *Proceedings of the British Academy*, 42, 109–132.
 Bok, Hilary (1999). *Freedom and Responsibility*. Princeton University Press.
 Capes, Justin (2019). What the Consequence Argument Is an Argument For. *Thought*, 8(1), 50–506.
 Chisholm, Roderick (1964). Human Freedom and the Self. The Lindley Lecture. Kansas University.

25. See discussion in Watson (2004: ch. 6).

- Elzein, Nadine and Tuomas Pernu (2017). Supervenient Freedom and the Free Will Deadlock. *Disputatio*, 9(45), 219–243.
- Frankfurt, Harry (1969). Alternate possibilities and moral responsibility. *The Journal of Philosophy*, 66(23), 829–839.
- Franklin, Cristopher (2015). Everyone Thinks that an Ability to Do Otherwise Is Necessary for Free Will and Moral Responsibility. *Philosophical Studies*, 172(8), 2091–2107.
- Grzankowski, Alex (2014). Can and the Consequence Argument. *Ratio*, 27(2), 173–189.
- Hale, Bob (2012). What is Absolute Necessity? *Philosophia Scientiae*, 16(2), 117–148.
- Horgan, Terence (1979). ‘Could’, Possible Worlds, and Moral Responsibility. *Southern Journal of Philosophy*, 17(3), 345–358.
- Ismael, Jenann (2016). *How Physics Makes Us Free*. Oxford University Press.
- Kaiserman, Alex (2022). Alternative Possibilities in Context. *Inquiry*, 65(10), 1308–1324.
- Kratzer, Angela (1977). What ‘Must’ and ‘Can’ Must and Can Mean. *Linguistics and Philosophy*, 1(3), 337–355.
- Lehrer, Keith (1976). ‘Can’ in Theory and Practice: A Possible Worlds Analysis. In Myles Brand and Douglas Walton (Eds.), *Action Theory* (241–270). Reidel.
- List, Christian (2014). Free Will, Determinism, and the Possibility of Doing Otherwise. *Noûs*, 48(1), 156–178.
- List, Christian (2019). *Why Free Will Is Real*. Harvard University Press.
- O’Connor, Timothy and Cristopher Franklin (2022). Free Will. In Edward N. Zalta and Uri Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition). <https://plato.stanford.edu/archives/win2022/entries/freewill>
- Russell, Paul (1988). Causation, Compulsion and Compatibilism. *American Philosophical Quarterly*, 25(4), 313–321.
- Schlick, Moritz (1939). *Problems in Ethics*. Trans. David Rynin. Prentice-Hall, Inc.
- van Inwagen, Peter (1975). The Incompatibility of Free Will and Determinism. *Philosophical Studies*, 27(3), 185–99.
- Watson, Gary (2004). *Agency and Answerability*. Oxford University Press.
- Weir, Ralph (2016). Relative Modality and the Ability to do Otherwise. *European Journal of Analytic Philosophy*, 12(1), 47–62.
- Whittle, Ann (2010). Dispositional Abilities. *Philosophers’ Imprint*, 10(12), 1–23.
- Wright, John (2022). Compatibilist Libertarianism: Why It Talks Past the Traditional Free Will Problem and Determinism Is Still a Worry. *Journal of the American Philosophical Association*, 8(4), 1–19.