The Dilemma Defense and Remaining Agnostic in the Right Way

Derek Lam

November 11, 2017

Abstract

The Frankfurt-style argument assumes determinism yet expects us to retain intuitive judgments about moral responsibility in particular scenarios. According to the Dilemma Defense, that's begging the question against incompatibilism. John Martin Fischer (2010) suggests that there's a new way to extract the compatibilist insight from the Frankfurt's cases with a slightly different argument, which I'll call the Irrelevance Argument. I'll examine two recent objections to the argument and show that they both fail. Then, I'll offer my own objection to Fischer's argument. Remaining agnostic about something isn't always theoretically innocent. I'll argue that Fischer's argument requires one to remain agnostic about how to analyze the ability to do otherwise in a very peculiar way that, despite its consistency, puts one in an objectionable dialectic position. I conclude that the Irrelevance Argument isn't a feasible way to revive the Frankfurt-style argument.

1 Fischer's Irrelevance Argument

It's well known that the compatibilist argument based on Frankfurt-style cases faces the Dilemma Defense (e.g., Goetz 2005). I assume familiarity with the Frankfurt's (1969) case about Black and Jones. If Black detects a prior sign that indicates that Jones is about to choose to vote for the Republican, Black would intervene with a device that causes Jones to choose to vote for the Democrat. It just so happens that Jones votes for Democrat on his own. Here's the dilemma. Either causal determinism is true or not. If it isn't, there won't be prior signs that Black can exploit to make sure that Jones absolutely cannot choose to do otherwise. Then, we don't have a case of a person who can't do otherwise but is still responsible for his action. But if causal determinism is true, it would be *part of the incompatibilists' view* that Jones isn't responsible for his action. It would then be begging the question to say that Jones is responsible for his action.

Conceding the force of the dilemma, Fischer (2010) suggests that Frankfurt's insight is better captured by a different argument, which I'll call the **Irrelevance Argument**. The Irrelevance Argument uses Frankfurt's case in a way that allows

Fischer to embrace the deterministic horn of the dilemma without begging the question. I interpret his argument as follows. ([1] - [2] are background conditions.)

> **[1]** Remain agnostic about whether determinism rules out Jones' ability to do otherwise.
>
> **[2]** Remain agnostic about whether Jones is responsible for his action.
>
> **Premise 1.** Determinism and the presence of Black's setup jointly rule out Jones's ability to do otherwise.
>
> **Premise 2.** The presence of Black's setup is irrelevant to whether Jones is responsible for his action.
>
> **Conclusion.** Even if Jones is not responsible for his action, that isn't because he lacks the ability to do otherwise.

Here's why the Irrelevance Argument seems valid. Given that we remain agnostic about whether determinism rules out Jones' ability to do otherwise (i.e., [1]), saying that determinism plus Black's setup does so *nonetheless* (i.e., Premise 1) is to say that Black's setup contributes to and hence at least partly explains the fact that Jones lacks the ability to do otherwise.[1]Then, the only way Black's setup can *still* be irrelevant to whether Jones is responsibility for his action (i.e., Premise 2) is for Jones's lack of ability to do otherwise to be irrelevant to whether he's responsible for his action. Hence, *if* Jones isn't responsible for his action (we remain agnostic about that according to [2]), that isn't because he lacks the ability to do otherwise.

What the Irrelevance Argument delivers isn't the outright compatibility of determinism and responsibility, but a slightly weaker compatibilist claim: even if determinism rules out responsibility, that isn't because it rules out alternate possibilities. The argument doesn't beg the question against incompatibilism. We assume determinism. But if we assume neither that Jones is responsible for his action nor that determinism doesn't rule out Jones's ability to do otherwise, the incompatibilists have no grounds for complaint. The Dilemma Defense appears to be circumvented.

The goal of this paper is to evaluate Fischer's Irrelevance Argument. I'll first examine two objections to the Irrelevance Argument. David Palmer (2014) challenges Premise 2 and Yishai Cohen (2017) questions the compatibility of Premise 1 and the agnostic condition [1]. I'll show that both objections fail. I'll then present my reason for rejecting the Irrelevance Argument. Traditional wisdom says that the fewer commitments an argument makes, the more forceful it is. I'll argue that this requires qualification. Remaining agnostic isn't always innocent. Depending

[1]Notice that Premise 1 does not say that determinism and the presence of Black's setup ***would have*** ruled out Jones's ability to do otherwise. It says that the two ***do*** jointly rule out Jones's ability. A subjunctive claim would not let us assert that Black's setup actually contributes something to Jones's inability to do otherwise.

on the dialectic context, remaining agnostic about certain things can carry theoretical burden that weakens one's argument. In particular, I'll argue that the agnostic condition [1] — despite being consistent with the Irrelevance Argument — puts one in an objectionable dialectic position.

2 Palmer on Premise 2

Premise 2 is rather standard for the Frankfurt-type cases. Black's setup, which remains unactivated, isn't part of the causal process that produces Jones's action. It's plausible to think that something that isn't part of the production of an action is irrelevant in determining who's responsible for the action. Premise 2 is the reason that Jones appears responsible for his action in Frankfurt's case.

Palmer (2014) questions Premise 2 by challenging the following principle:

> **(IP-W)** If a fact is irrelevant to the causal explanation of a person's action, then that fact is irrelevant to whether or not that person is morally responsible for her action. (3856)

If IP-W is false, Premise 2 is unmotivated. Palmer offers the following counterexample to IP-W:

> Suppose that Jones finds himself with an irresistible desire to decide to break his promise. [...] Jones is a moral person. So, upon finding himself with this desire, Jones tries his very best to resist it. However, despite trying his very best to resist the desire, he eventually succumbs to it and decides to break his promise.
>
> Now, the fact that Jones tried his very best to resist the desire to decide to break his promise is plausibly irrelevant to the causal explanation of his deciding to break his promise. After all, he decided to break his promise in spite of this effort, not because of it [...]. But the fact that he tried his best to resist the desire to decide to break his promise is plausibly relevant to whether or not he is morally responsible for deciding to break his promise: [that] is a reason why Jones is not blameworthy (at all) for deciding to break his promise. (3857)

This isn't a successful counterexample. It's noteworthy that a discussion about moral responsibility can be about two different issues. On the one hand, we can be interested in whether an agent is morally responsible for the action in the sense that we want to know whether the agent is *accountable* for or, in other words, *owns* the action so that moral evaluation is *applicable* at all. Let's call moral responsibility in this sense **ownership responsibility**. To say that a person is morally responsible for an action in this sense doesn't yet involve making any normative judgment. To say that one is accountable for doing X isn't saying anything about

whether one ought or ought not have done X yet. On the other hand, by inquiring into a person's moral responsibility for an action, we may be interested in knowing *how exactly* a person is responsible for the action — is that person blameworthy, praiseworthy, despicable, noble, etc. for doing that action. Let's call this a person's **evaluative responsibility** for an action. Unlike ownership responsibility, spelling out a person's evaluative responsibility involves making explicit *normative* judgements. I take my notion of ownership responsibility to be similar to Watson's (1996) notion of *accountability* or Pereboom's (2001; 2014) notion of *basic desert*.[2]

Debates about free will are about ownership responsibility. We want to understand, for example, if causal determinism is true, whether the actions we perform can still be considered ours (or: whether we're still accountable for any actions). On its own, this isn't a normative question.[3]Despite the adjective "moral", for example, when Frankfurt explores the compatibility of determinism and *moral* responsibility, he's only interested in the compatibility of determinism and our *ownership* of the actions we perform. It matters that Jones is account-*able* for his action. But it never matters whether Jones is to be blamed, praised, admired, or condemned for voting the way he did. Actual normative evaluation is irrelevant. Questions about evaluative responsibility are for *normative ethics* to settle. Something that is relevant for normative evaluations need not be relevant for determining whether a person is to be held accountable for an action in the first place. The former is an ethical issue, the latter a metaphysical (or meta-ethical) issue. Given the context, "moral responsibility" in IP-W should be read as referring to a person's *ownership* responsibility.

Now that we're clear how IP-W is meant to be understood, whether Palmer's example is a genuine counterexample crucially depends on whether it's a case in which a fact isn't relevant to the causal explanation of an action but remains relevant to deciding whether the agent has *ownership responsibility* for the action. Palmer's example fails because it elicits our intuition about evaluative responsibility to challenge a principle about ownership responsibility.

Here's an argument for thinking that Palmer's counterexample is about evaluation instead of ownership. Let's assume that ownership responsibility is a precondition for evaluative responsibility. That is, actions don't lead to ethical evaluations of ourselves if those actions aren't genuinely ours to begin with. If having tried to resist an irresistible desire to do X is relevant to whether a person has ownership responsibility as Palmer believes, having tried to resist to follow the irresistible desire to do X would invalidate ***any*** kind of evaluative responsibility (being blameworthy, being praiseworthy, being admirable, being a pervert, etc. for doing X). But the consequence is false.

[2]I don't claim that the three notions are exactly the same. But they are similar to the extent that making claims about ownership responsibility, claims about accountability, or claims about basic desert isn't to make any specific *normative* judgment. All I need for my subsequent argument to work is the idea that the notion of moral responsibility we care about in this context is a non-normative one.

[3]Also see Alfred Mele (1995: 3-4) for the similar idea. Mele argues that the notion of autonomy or free will we care about can be instantiated by organisms/gods that morality doesn't apply.

Consider the following two cases:

> [**Praiseworthiness**] Emily has a gentle soul. She cannot help but give a beggar all her change whenever she walks by one. Emily isn't a saint though; there is still selfishness in her. Walking past a beggar in a hot summer day, she tries her best not to give the beggar her change so she can get herself some nice gelato. But her resistance is futile. She can't help but give the beggar all her change.
>
> [**Perversity**] Peeping Tom has an irresistible desire to take a peep at Lady Godiva. He still has a little bit of decency in him left that he tries his very best not to decide to peep. But his resistance is futile. He can't help but peeps.

Had Emily not tried to resist her generosity, and had Tom not tried to resist peeping, Emily would have been *more* praiseworthy for her deed and Tom would have been *less* perverse. *Nonetheless*, we wouldn't deny that, as things stand in [Praiseworthiness] and [Perversity], Emily is still praiseworthy and peeping Tom is still perverse.[4]

I'm willing to grant Palmer the point that, in the scenario he describes, Jones isn't blameworthy *at all* for deciding to break his promise given that he has put in a fight. The question remains: does his futile attempt have this effect because it's relevant to determining whether the decision to break the promise is Jones' *own* decision (i.e., ownership responsibility)? Or does it have this effect because it's a factor in our *normative* reasoning about blameworthiness specifically (i.e., evaluative responsibility)?

It's hard to tell if we focus on Palmer's case alone. But by contrasting his case with [Praiseworthiness] and [Perversity], we have a compelling reason for thinking that trying to resist an irresistible desire to do X isn't relevant to undermining a person's ownership responsibility for doing X because, otherwise, Emily wouldn't have been praiseworthy *at all* for having tried to resist her generous decision and peeping Tom wouldn't have been a pervert *at all* for having resisted his objectionable urge. We should therefore conclude that the fact that Jones' attempt to resist lifts his blameworthiness completely has something to do with blameworthiness as *a specific kind of ethical evaluation*; it has nothing to do with the *metaphysics* of action ownership. Palmer fails to offer a counterexample to IP-W once it's properly understood.

Thus, Premise 2 remains intact. At least, Palmer hasn't offered any compelling reason for us to think otherwise.

[4] I present two cases, one of *positive* evaluation and one of *negative* evaluation, in order to preempt responses based on Susan Wolf's (1980) idea that ownership responsibility is disjunctive, depending on the moral status of the action.

3 Cohen on Premise 1

More recently, Cohen (2017) presents an interesting objection to the Irrelevance Argument by focusing on Premise 1 instead. He presents the Irrelevance Argument, which he calls "Fischer's Improved Argument", in a more expansive way than I do. But his point can be translated succinctly to apply to the Irrelevance Argument as I present it. On my reading, his objection to the Irrelevance Argument consists of two steps. First, he argues that Premise 1 can be interpreted in two different ways: either in terms of entailment or in terms of explanation:

> **Premise 1*.** Determinism and the presence of Black's setup jointly *entail* that Jones lacks the ability to do otherwise.
>
> **Premise 1**.** Determinism and the presence of Black's setup jointly *explain* why Jones lacks the ability to do otherwise.

Cohen argues that, whereas the argument wouldn't be valid if we interpret the first premise as Premise 1*, we can't know that the argument is sound if we interpret the first premise as Premise 1**.

Let's consider Premise 1* first. If determinism entails Jones' inability to do otherwise, determinism in conjunction with anything would have the same entailment due to the monotonicity of deduction. That is, if determinism alone entails Jones' inability to do otherwise, Premise 1* would be trivially true. That undermines the thought that Black's setup contributes to and is hence required for Jones' inability to do otherwise in the context of the thought experiment. As I have illustrated, that Black's setup contributed to Jones' inability to do otherwise is crucial to the validity of the argument. If Black's involvement isn't in fact relevant to Jones' inability to do otherwise, then the fact that Black's setup is irrelevant to Jones' responsibility tells us nothing about the relation between Jones' lack of ability to do otherwise and his lack of responsibility. So, step one: the Irrelevance Argument must be interpreted in terms of explanation (i.e., Premise 1**); since explanation isn't monotonic like entailment, hopefully, Premise 1** can avoid the problem.

The second step of Cohen's objection says that [1] is in conflict with the *assertion* of Premise 1**. Hence, if the agnostic condition [1] is in place, we aren't in a position to offer the Irrelevance Argument as a sound argument. Here's Cohen's reasoning. If determinism alone ruled out Jones' ability to do otherwise, that would rule out other explanations for Jones's inability to do otherwise. In particular, that rules out explanations that appeals to Black's setup. So, with [1], one cannot assert that Black's setup contributes to explaining Jones's inability. This is a problem for Fischer that can be fleshed out in two ways.

Here's one way to look at it. If one believes that, for P and Q to *jointly* explain R, P and Q must individually contribute to explaining R, then the idea that [1] forces us to refrain from asserting that Black's setup contributes explanatorily to Jones' situation implies that we cannot assert Premise 1** as long as [1] is in place. Here's another way to articulate the point. If Black's setup doesn't contribute to

explaining Jones' inability, then what I said about Premise 1* applies again: if Black's involvement doesn't contribute to Jones' inability to do otherwise, then the fact that Black's setup is irrelevant to Jones' responsibility tells us nothing about the relation between Jones' lack of ability to do otherwise and his lack of responsibility. The conclusion doesn't follow from Premise 1** and Premise 2.

I agree with the first step of Cohen's objection (though I believe Fischer is well aware of that, which is why he emphasizes that Black is also *necessary* for ruling out Jones' ability). Let's focus on step two.

What does the heavy lifting in step two is some kind of anti-overdetermination principle about explanation: if P alone explains Q, there can't be an R that also explains Q (assuming P $\neq$ R). More specifically, Cohen employs a narrower anti-overdetermination claim: if P alone explains Q *by entailment*, there can't be an R that also explains Q.[5]If we hold on to [1] and assert Premise 1**, we basically embrace the epistemic possibility that, while determinism alone explains Jones' inability by entailment, determinism plus Black's setup also explains Jones' inability. (Note: determinism $\neq$ determinism plus Black's setup.) That would be an epistemic possibility of overdetermination. If Cohen's narrow anti-overdetermination claim is right, we shouldn't accept this epistemic possibility.

Step two doesn't work. On the topic of overdetermination and in response to the Dilemma Defense, Fischer argues that, even if determinism in fact explains Jones' inability to do otherwise, unless we assume that one explanation must "crowd out" other explanations, there is nothing that prevents Black's setup from being a part of an explanation for Jones' inability as well. Admittedly, this remark alone doesn't achieve much. It establishes neither the possibility of overdetermination generally nor the plausibility of overdetermination in Jones' case specifically. Fortunately, for the Irrelevance Argument to circumvent Cohen's worry, ***agnosticism*** about overdetermination is enough. We add a third agnostic condition to the Irrelevance Argument:

> **[3]** Remain agnostic about whether, *if* Jones lacked the ability to do otherwise, there is an *overdetermination of explanations* for the fact that he lacked that ability.

On my reading, that's exactly the point of Fischer's remark: it's not obvious at all that overdetermination can't be present in Jones' case.

If we remain agnostic about whether determinism's entailment of Jones' inability (*if* there were such an entailment relation) rules out other explanations, and if we also remain agnostic about whether there is such an entailment relation, then it becomes coherent again to say that Black's setup contributes to an explanation of Jones' inability. Certainly, if Black's setup contributes something, then it can't

[5]Cohen phrases this claim differently by adding a precondition for partial explanation that there must be no overdetermination (2017: 132). I think what I say here captures the essence of his objection.

both be the case that there is an entailment relation and that there are no overdetermination. At most one of them can be true. But we can remain agnostic which of them is true and, hence, remain agnostic about both.

The validity of the Irrelevance Argument depends on the thought that Black's setup contributes something to Jones' inability. Of course, the coherence of having that thought isn't enough. It has to be plausible for the Irrelevance Argument to be valid. But, as I've explained in section 1, it's plausible when the agnostic condition [1] and Premise 1 (or Premise 1**) are put together: if I don't know whether determinism alone explains Jones' situation but I say that determinism plus Black's setup does, I'm basically saying that Black's device is contributing at least something to explain Jones' situation.

In response to Fischer's remark about overdetermination, Cohen may reply that, instead of arguing against overdetermination generally, he only defends the narrower claim that an explanation *by entailment* "crowds out" other explanations.[6]But weakening his anti-overdetermination stance in an ad hoc manner wouldn't help him fence off Fischer's agnostic remark about overdetermination and hence wouldn't help Cohen strengthen his dialectic position. Once he gives up defending anti-overdetermination generally, the ground for the narrower claim becomes shaky. Cohen offers two examples where one explanation by entailment crowds out other explanations. But notice that overdetermination sympathizers don't think that explanations never crowd out each other. So, it's not clear individual cases like that alone carry much weight. Furthermore, it's instructive to note that the advocates of the Irrelevance Argument doesn't need to affirm overdetermination in Jones' case, they only need to remain agnostic. It's Cohen, who must justify a positive anti-overdetermination claim. The dialectic favors Fischer.

4 Just the Right Amount of Agnosticism

The Irrelevance Argument is valid. And I've defended both premises from objections by Palmer and Cohen. Does that mean Fischer has successfully revived a compatibilist argument based on the Frankfurt-style cases in response to the Dilemma Defense?

Like Cohen, I believe the relation between Premise 1 and [1] calls for a closer scrutiny. Unlike Cohen, I don't think there is any inconsistency in asserting Premise 1 with the background condition [1] in play. Nonetheless, as I'll argue, insofar as we accept Premise 1, remaining agnostic as [1] requires introduces certain theoretical burden that is dialectically objectionable.

Let's consider [1] first. Remaining agnostic about something isn't as theoretically innocent as it's commonly assumed to be. How can it be reasonable to remain agnostic about whether determinism rules out Jones' ability to do otherwise? To the libertarians, for instance, the ability to do otherwise *conceptually*

[6]Cohen writes, "Notice that I have not defended (and need not defend) the position that an overdetermination of explanation [. . .] is impossible." (2017: 135)

consists of some kind of indeterminacy. So, determinism and Jones' ability to do otherwise is plainly inconsistent. If the Irrelevance Argument can't work unless we remain agnostic about a plain inconsistency, then so much for a better argument for compatibilism.

Fortunately, the indeterminist analysis of "the ability to do otherwise" isn't the only game in town. The Humean conditional analysis of "the ability to do otherwise", despite its many problems, famously doesn't conflict with determinism. The same applies to Vihvelin's (2004; 2013) dispositional analysis. It's clear that the compatibilists can embrace neither Hume's nor Vihvelin's analyses as part of the Irrelevance Argument because the argument would basically be useless if it required us to first accept a compatibilist analysis of the ability to do otherwise. But since there are analyses of "the ability to do otherwise" that don't conflict with determinism, we can at least remain agnostic about which analysis — including the incompatibilist analyses — is correct. Thus, we can reasonably remain agnostic about whether determinism rules out Jones's ability to do otherwise, i.e., [1].

Let's now turn our attention to Premise 1. Fischer explains Premise 1's plausibility in the following way:

> Why exactly do I say that causal determinism plus Black rules out alternative possibilities, when I am not here supposing that mere causal determinism does not? Well, it is supposed to work as follows. Black knows that, given that Jones has exhibited the Democratic sign at t1, he need not intervene at all since Jones is going to vote for the Democrat. But, given our assumptions, Jones can exhibit the Republican sign at t1. But Black will be there monitoring the situation, and if he were to see the Republican sign at t1, then he would immediately zap Jones's brain and thereby prevent Jones from choosing to vote for McCain at t2 (or voting for McCain at t3). Without Black, there is nothing in the example that rules out Jones's power to choose and do otherwise; **but with Black** (together with causal determinism), **we get the result that Jones cannot choose at t2 to vote for McCain** (and cannot so vote at t3). (2010: 329; emphasis added)

It's unclear why Fischer thinks he's justified for accepting the claim I highlighted. It's at least not obviously plausible by the light of the conditional analysis of the ability to do otherwise. Surely, had Jones ever been about to choose otherwise, he still wouldn't have chosen otherwise because of Black's intervention. Nonetheless, he arguably would have voted differently ***had Jones gotten to choose*** to vote differently. That's because, given the assumption about determinism and Black's device, the closest possible world(s) at which Jones gets to choose differently despite showing prior sign that he is about to choose differently at some point are worlds at which either Black's device malfunctions or Black is distracted — presumably Jones ends up succeeding in voting differently at such worlds. That, according to the conditional analysis, just is the ability to do otherwise, which isn't removed by Black's setup.

Fischer's explanation also wouldn't be plausible by the light of Vihvelin's dispositionist analysis of the ability to do otherwise. Crucial to making Vihvelin's analysis of the ability to do otherwise compatibilist is the idea that a disposition can be ***masked*** (by causal determinism) without being taken away.[7]By taking good care of an expensive vase, I've masked the fragility of the vase: even if someone knocked it off the table, I would be there to catch it. My presence prevents the vase's fragility from manifesting. Intuitively, however, the vase is still fragile; the disposition is still there. As Vihvelin points out, in order to determine whether something has ability to do something, we need test-cases where we set aside "extrinsic masks":

> S has the narrow ability at time t to do R in response to the stimulus of S's trying to do R iff, for some intrinsic property B that S has at t, and for some time t' after t, if S were in a test-case at t and S tried to do R and S retained property B until time t', then in a suitable proportion of these cases, S's trying to do R and S's having of B would be an S-complete cause of S's doing R.
>
> [To be in a test-case is to be] in surroundings where the extrinsic enablers (e.g., a bicycle and a place to ride it) for the ability are in place and where there are **no extrinsic masks** (e.g., bicycle bullies lurking in the background) to the exercise of the ability [...] (2013:187; emphasis added)

In Jones' case, Black's standing by to intervene is exactly an *extrinsic masks* that need to be set aside. If we understand the ability to do otherwise as the disposition to do otherwise, then, just like my presence merely masks the fragility of the vase, Black's setup only masks Jones' ability to do otherwise. *If* Black's device were not in place, Jones would very likely be *doing* otherwise if he were to *choose* or try to do otherwise. Hence, Jones still has the ability to do otherwise. Committing ourselves to Vihvelin's dispositional analysis would not allow us to embrace Premise 1.

Here's a first pass of my concern about the Irrelevance Argument. If I'm right that Fischer can reasonably remain agnostic according to [1] only because he can remain non-committal on the options for analyzing the ability to do otherwise, asserting Premise 1 becomes tricky. That's because the assertion of Premise 1 doesn't permit one to remain neutral with respect to the analysis of the ability to do otherwise. The Irrelevance Argument is therefore self-undermining, as Cohen says. But the self-undermining feature that I'm alluding to has nothing to do with overdetermination of explanation and, hence, introducing [3] wouldn't help.

That's too quick. Perhaps holding on to [1] only requires one to remain neutral with respect to the analysis of the ability to do otherwise to a certain extent and asserting Premise 1 only requires one to be selective about the analysis of the

[7]See also Michael Fara (2008).

ability to do otherwise to a certain extent. Perhaps holding onto [1] and asserting Premise 1 isn't really inconsistent once we are clear to what extent we should be neutral and to what extent we should be selective about the analysis of the ability to do otherwise.

Based on what we've seen, we can tell that the Irrelevance Argument requires us to be able to remain open to a very particular range of options with respect to analyzing the ability to do otherwise. Specifically, his argument requires us to leave room for a ***candidate*** analysis that cannot be ruled out by determinism but can be ruled out by determinism plus Black's setup.

It's noteworthy that none of the analyses proposed so far has this feature. For the obvious reason that I've mentioned, no incompatibilist proposals can fit the description. None of the actual compatibilist proposals fit the description either (i.e., proposals that promise to make determinism compatible with the ability to do otherwise, which may or may not be relevant to free will).

Earlier I argued that the dispositional analysis doesn't fit the bill because Jones's ability to do otherwise wouldn't be ruled out by determinism plus Black. One might be sympathetic to the dispositional account but think that having the ability to do otherwise requires more than just having opposing powers. For example, Randolph Clarke & Thomas Reed (2015) argue that having opposing powers alone is not enough for *free will* as long as whether the conditions required for *exercising* said powers obtain or not *is not up to the agent*. Our present concern isn't free will but the ability to do otherwise, which doesn't have to be sufficient for free will. Nonetheless, one might argue that a similar constraint must be built into our analysis of, not only free will, but also the ability to do otherwise. If so, since the presence of determinism plus Black is not up to Jones, *the beefed up dispositional ability to do otherwise* would be ruled out by determinism plus Black. Does that mean there is a compatibilist analysis that fits our description after all?

Unfortunately, I think the beefed up dispositional analysis is *incompatibilist*. If the presence of determinism plus Black isn't up to Jones, so are the things that happened a thousand years ago and the laws of nature. The dispositional analysis can be compatibilist because it allows factors that are not up to an agent (e.g., events in the remote past) to prevent the manifestation of her ability to do otherwise without eliminating that ability. By requiring the ability to do otherwise to be incompatible with such factors, the dispositional analysis is beefed up in a way that it's no longer compatibilist.[8] Hence, we still don't have an analysis that makes determinism not rule out Jones's ability to do otherwise but makes determinism plus Black rule out Jones's ability to do otherwise.[9]

[8]This is basically van Inwagen's Consequence Argument. And I'm not sure what to make of Clarke & Reed's remark when they say they are not sure whether the beefed up dispositional analysis is compatibilist because they are not sure whether there can be good response to the Consequence Argument applied to this beefed up dispositional analysis. If one doesn't have a good objection to the Consequence Argument here, one should then accept, I think, as I do, that the beefed up dispositional analysis makes the ability to do otherwise incompatible with determinism.

[9]I'm grateful to Stephen Kearns and Michael McKenna for pressing me very hard on this

So, oddly enough, embracing the Irrelevance Argument requires us to accept that there is a special proposal *out there* for analyzing the ability to do otherwise that isn't on the table yet. Let's call this yet-to-be-articulated proposal **Analysis X**.

I'm not pessimistic about the possibility of introducing Analysis X to complete the Irrelevant Argument. But it's worth asking: what move can the compatibilists *aim for* here to complete the argument? To introduce Analysis X as a legitimate option, Analysis X should, of course, be properly articulated and independently motivated. Suppose we seek motivation for Analysis X in an argument R. Apparently, there are two possible outcomes one may aim for: (1) present R as a convincing reason for accepting Analysis X; (2) present R as an argument that appears to be convincing but in fact falls short of being a convincing reason for Analysis X.

For a reason to be convincing, it doesn't have to be *conclusive* or *entail* whatever it's a reason for. A defeasible yet convincing reason for accepting Analysis X is just one, given which, for all we know (including, for instance, everything we know about other available analyses and arguments with respect to the (in)compatibilism debate *up to this point*), we **should** (or, if you are a permissivist: **are rationally permitted to**) accept Analysis X even though *future* evidence may override that demand (or: permission). The difficulty for the advocates of the Irrelevance Argument is that they *shouldn't* aim for (1) and cannot aim for (2).

Note that Frankfurt's targets are people who think that determinism rules out free will or moral responsibility *by ruling out* the ability to do otherwise. An independent and convincing reason for Analysis X is a convincing reason for denying that determinism can rule out either free will or moral responsibility by ruling out the ability to do otherwise.[10]Successfully doing so renders the attempt to present the Irrelevance Argument as an argument against incompatibilism superfluous. That's why the advocates of the Irrelevance Argument shouldn't aim for (1).

Korsgaard (2009) argues that, although we can intend to do something but *happen to* fail, we can't intend to do something unsuccessfully. The cases where we say a person does X unsuccessfully on purpose are cases of the person *pretending* to do X. Suppose I'm in a restaurant with friends. Let's further assume that I'm a cheap and manipulative person who wants to appear generous. So, I declare to pay for everyone, hoping that one of my friends would feel bad and stop me from doing so. That way, I can appear generous without actually having to pay for everyone. In this case, it's not my intention to pay unsuccessfully because I don't intend to pay at all. If I don't intend to do X at all, I don't intend to do X in any particular way. Instead, a proper description of what I intend to do is *pretending* to pay for everyone. It's part of the metaphysical nature of intentional action that there can't be intentional failures.

issue.

[10]Compatibilism wins either we *ought* to accept a compatibilist analysis of the ability to do otherwise or we are rationally *permitted* to accept an analysis that is compatibilist. So, the permissivism debate does not make a difference here.

Let's consider (2) in light of Korsgaard's claim. Intuitively, a reason/argument is *meant* to be convincing. To have an unconvincing argument is to have no argument at all. Offering an unconvincing reason to introduce a proposal is an unsuccessful attempt to motivate a proposal. Since it's impossible to intentionally do something unsuccessfully, it's impossible to aim for (2).

If what I've said about the dialectic peculiarity of [1] and Premise 1 is right, advocating the Irrelevance Argument puts one in a dialectically awkward position: somehow we need to aim to offer a convincing reason for a *proposal* that fits the profile of Analysis X — so we have the right set of open options to remain agnostic about the ability to do otherwise in the way the Irrelevance Argument requires[11]— while simultaneously hope that our argument *will eventually* fall short of being convincing — so that the Irrelevance Argument won't end up superfluous.

It might not be *inconsistent* to aim to argue for something with the hope that the argument will fail eventually. But it's dialectically irrational.[12]Hence, I conclude that, given that we don't already have something like Analysis X (so that it must be introduced as a legitimate option), there simply isn't a reasonable way for the compatibilists to *pursue* the Irrelevance Argument. The problem lies not in the argument itself but in the irrational condition a compatibilist needs to put herself through to remain agnostic in just the right way so that she may *use* the Irrelevance Argument in the current dialectic situation.

References

Alfred, R. M. (1995). *Autonomous Agents: From Self Control to Autonomy*. Oxford University Press.

Clarke, R., & Reed, T. (2015). Free Will and Agential Powers. *Oxford Studies in Agency and Moral Responsibility*, *3*, 6–33.

Cohen, Y. (2017). Fischer's Deterministic Frankfurt-Style Argument. *Erkenntnis*, *82*(1), 121–140.

[11]One might wonder whether I'm holding a double standard by including the Humean analysis as a candidate analysis for the ability to do otherwise even though we don't have convincing reason for accepting it (anymore) while requiring the advocates for the Irrelevance Argument to aim for providing a convincing reason for Analysis X in order to introduce Analysis X as a candidate analysis. It isn't double standard. Dialectic considerations are inherently sensitive to the ***history*** of the discussion. The condition under which it's reasonable to ***continue*** to treat a view as a candidate (that includes not ruling out that some refined version of the view may prove to be correct) need not be the same as the condition under which it's reasonable to ***introduce*** a view as a candidate in the first place. In particular, I find it plausible to think that the bar for introducing a new candidate is higher than keeping a previously introduced candidate alive. When the Humean analysis was introduced, it was with new insight and defeasible yet convincing reasons. The fact that those reasons aren't considered forceful isn't in conflict with the belief that the view is roughly in the right direction and some refined version of it may prove to be correct someday.

[12]I find this claim rather plausible. But I confess that I don't really have an argument for it.

Fara, M. (2008). Masked Abilities and Compatibilism. *Mind*, *117*(468), 843–865.

Fischer, J. M. (2010). The Frankfurt Cases: The Moral of the Stories. *Philosophical Review*, *119*(3), 315–336.

Frankfurt, H. (1969). Moral Responsibility and the Principle of Alternative Possibilities. *Journal of Philosophy*, *66*, 829–839.

Goetz, S. C. (2005). Frankfurt-Style Counterexamples and Begging the Question. *Midwest Studies in Philosophy*, *29*(1), 83–105.

Korsgaard, C. M. (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press.

Palmer, D. (2014). Deterministic Frankfurt Cases. *Synthese*, *191*(16), 3847–3864.

Pereboom, D. (2001). *Living Without Free Will*. Cambridge University Press.

Pereboom, D. (2014). *Free Will, Agency, and Meaning in Life*. Oxford University Press.

Vihvelin, K. (2004). Free Will Demystified: A Dispositional Account. *Philosophical Topics*, *32*(1/2), 427–450.

Vihvelin, K. (2013). *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*. Oxford University Press.

Watson, G. (1996). Two Faces of Responsibility. *Philosophical Topics*, *24*(2), 227–248.

Wolf, S. (1980). Asymmetrical Freedom. *Journal of Philosophy*, *77*(March), 151–66.