

Penultimate draft. Forthcoming in *Philosophical Explorations*.

Do intuitions about Frankfurt-style cases rest on an internalist prejudice?

Florian Cova

Campus Biotech, CISA – University of Geneva

Case Postale 60, CH - 1211 Genève 20

florian.cova@gmail.com

Hichem Naar

Centre for Research in Ethics/Dept of Philosophy – University of Montreal

Pavillon 2910, boul. Édouard-Montpetit - Montréal QC H3C 3J7

hm.naar@gmail.com

‘Frankfurt-style cases’ (FSCs) are widely considered as having refuted the Principle of Alternate Possibilities (PAP) by presenting cases in which an agent is morally responsible even if he could not have done otherwise. However, Neil Levy has recently argued that FSCs fail because (i) our intuitions about cases involving counterfactual interveners are inconsistent (we accept that the mere presence of counterfactual interveners is enough to make us gain but not lose responsibility-underwriting capacities), and (ii) this inconsistency is best explained by the fact that our intuitions about such cases are grounded in an internalist prejudice about mental states and capacities. In response to this challenge, we argue (i) that there is no inconsistency in our intuitions about cases involving counterfactual interveners, as soon as one draws the comparison properly, and (ii) that intuitions about such cases do not rest on an internalist prejudice, but on a more basic distinction between two kinds of dispositions. Additionally, we discuss some methodological issues that arise when comparing intuitions about thought experiments and end with a discussion of the implications of our argument for the reliability of intuitions about FSCs.

Keywords: alternate possibilities; capacities; dispositions; Frankfurt cases; free will; moral responsibility

For years, the philosophical debate about the compatibility of moral responsibility with determinism was dominated by the Principle of Alternate Possibilities (PAP):

(PAP) Persons are morally responsible for what they have done only if they could have done otherwise.

Under a certain interpretation of ‘could have done otherwise’, one can deduce from PAP the incompatibility of moral responsibility with determinism for, under that interpretation, an agent in a deterministic world cannot do otherwise. A heated debate ensued on whether this interpretation of ‘could have done otherwise’ was the one relevant to moral responsibility. Compatibilist philosophers (who argue for the compatibility of moral responsibility with determinism) denied that it was, while incompatibilists (who argue for their incompatibility) argued that it was not.

Many people consider that this ‘dialectical stalemate’ (Fischer, 1994) was broken in 1969 with the publication of Harry Frankfurt’s seminal paper ‘Alternate Possibilities and Moral Responsibility’ (Frankfurt, 1969). In this paper, Frankfurt explained how to design thought experiments in which an agent is morally responsible *even if* he could not have done otherwise. This is achieved by constructing cases in which an agent acts entirely *on his own*, but in which a *counterfactual intervener* (henceforth: CI) would have forced him to act the same way if he hadn’t. These ‘Frankfurt-style cases’ (FSC) are widely held to constitute counter-examples to the PAP by suggesting that the ability to do otherwise might not be necessary for moral responsibility.

However, not everyone has been convinced by FSCs’ ability to defuse the PAP. Some have argued that agents in FSCs still have alternate possibilities (e.g. van Inwagen, 1983). Others have argued that FSCs presuppose determinism and hence that claiming that agents in FSCs are morally responsible only begs the question against incompatibilism (e.g. Ginet, 1996). More recently, Neil Levy (2008) has argued that FSCs fail because they suppose that

the presence of a CI has no impact on the agent's responsibility-underwriting capacities. Indeed, Levy provides examples suggesting that the mere presence of a CI *can* be enough to make an agent *gain* responsibility-underwriting capacities. If so, FSCs' presupposition that the mere presence of a CI cannot alter an agent's responsibility-underwriting capacities might be nothing more than an unwarranted 'internalist' prejudice according to which mental states and capacities are wholly internal to subjects.

In this paper, we argue against Levy's conclusion: people's intuitions about FSCs do not rest on some implicit internalist commitments. Rather, they rest on an intuitive distinction between two kinds of abilities, which itself rests on a more basic distinction between two kinds of dispositions. Thus, the intuition that responsibility-underwriting abilities can survive the presence of a counterfactual intervener, even if other abilities can be lost very similar conditions, has nothing to do with internalism about the location of mental states (henceforth 'internalism'). An upshot of the discussion is that it is not clear that Levy's argument succeeds in explaining away our intuitions about FSCs.

1. Levy's objection to Frankfurt counter-arguments

Here is a standard FSC (Pereboom, 2001):

Republican Voting: When she enters the voting booth, Connie has not yet made up her mind whether to vote Democratic or Republican. Unbeknownst to her, an evil but gifted neuroscientist monitors her neural states using a computer chip he has implanted in her brain. The chip gives the neuroscientist the power to intervene in Connie's neural processes. Connie is psychologically constituted so that it is a necessary (though not sufficient) condition of her voting Democratic that she thinks deeply about a certain policy; were Connie to think deeply about this policy, the neuroscientist would intervene, causing her to choose to vote Republican. But Connie does not think deeply about the policy and

proceeds to vote Republican on her own. The neuroscientist and his device play no role in bringing about Connie's vote.

This case is supposed to show that Connie can freely vote Republican (and be morally responsible for her vote) even if she couldn't have done otherwise (because of the neuroscientist's chip). Cova (2014) distinguishes two different arguments that could be drawn from such cases to reject PAP: the *direct* and *indirect* Frankfurt counter-arguments.

The *indirect Frankfurt counter-argument* (IFC) is supposed to go like this: imagine the same case as *Republican Voting*, except that there is no neuroscientist. In this case – let's call it *Normal Voting* – Connie acts normally, so that everyone would agree that she's free and morally responsible. Now, let's compare *Normal Voting* and *Republican Voting*: in *Republican Voting*, Connie acts exactly the same way as in *Normal Voting*, since the neuroscientist does not act. Surely, the mere presence of the neuroscientist cannot have an impact on the way Connie acts and the way she came to act this way. But, if the mere presence of the neuroscientist does not change anything as far as Connie is concerned, we must conclude that Connie is responsible in *Republican Voting*, for she was responsible in *Normal Voting*. However, Connie could not have done otherwise. Therefore, Connie is morally responsible *even if* she could not have done otherwise, and this counts as a counter-example to the PAP.

More generally, IFC can be summarized in the following way:

The indirect Frankfurt counter-argument:

1. The agent is morally responsible in the *Normal* case.
2. Since the only difference between the *Normal* case and the *Frankfurt* case is the presence of a CI, and since this CI does not intervene in the *Frankfurt* case, there is no

significant difference between the *Normal* and *Frankfurt* cases as far as the agent's moral responsibility is concerned.

3. Thus, the agent is morally responsible in the *Frankfurt* case.
4. However, due to the presence of the CI, the agent could not have done otherwise in the *Frankfurt* case.
5. Therefore, the PAP is false, since the agent in the *Frankfurt* case is morally responsible even though he could not have done otherwise.

Levy rejects this counter-argument by casting doubt upon premise 2. For Levy, it is not clear that the presence of a CI makes no difference, for it might be that his presence is enough to make the agent lose some responsibility-underwriting capacities. Thus, the inference from the fact that the agent is responsible in the *Normal* case to her being responsible in the *Frankfurt* case is not warranted.

How does Levy show that the mere presence of a CI could make an agent lose responsibility-underwriting capacities? Actually, he does not show directly that the presence of a CI can cause the *loss* of responsibility-underwriting capacities. Rather, his strategy consists in introducing new cases, the 'Frankfurt-style enabling cases' (FECs), in which the presence of a CI is enough to make an agent *gain* responsibility-underwriting capacities, and then argue that if the presence of CI is enough to make an agent *gain* responsibility-underwriting capacities, there is no reason to think that it cannot also make an agent *lose* responsibility-underwriting capacities.

Here is an example of FEC:

Phobia: Jillian is walking along the beach when she notices a child drowning. Jillian is a good swimmer, but she is pathologically afraid of deep water. She is so constituted that her phobia would prevent her from rescuing the child were she to attempt to; she would be overcome by feelings of panic.

Nevertheless, she is capable of trying to rescue the child, and she knows that she is capable of trying. Indeed, though she knows that she has the phobia, she does not know just how powerful it is; she thinks (wrongly) that she could affect the rescue. Unbeknownst to Jillian, a good-hearted neurosurgeon has implanted her with a chip with which he monitors Jillian's neural states, and through which he can intervene if he desires to. Should Jillian decide (on her own) to rescue the child, the neurosurgeon will intervene to dampen her fear; she will not panic and will succeed, despite her anxiety, in swimming out to the child and rescuing her.

Imagine that Jillian does not try to save the child and the child drowns. Presumably, she is responsible for the child's death, for she could have saved him. But what if she had let the child drown in a similar scenario *without* the presence of the neurosurgeon? It seems that, in this case, she would not have been responsible for the child's death. According to Levy, this means that the mere presence of the neurosurgeon is enough to make Jillian morally responsible for the child's death, that is, to make her *gain* responsibility-underwriting capacities.

But if the mere presence of a CI is enough to alter an agent's responsibility-underwriting capacities, then we are no longer entitled to take premise 2 for granted, no matter how intuitive it seems to us. On the contrary, we should be cautious of the intuition that CIs cannot make any difference to an agent's responsibility-underwriting abilities, for this intuition may rest on an internalist prejudice. As Levy puts it,

we cannot accept that the agents who feature in FSCs actually possess the capacities needed to justifiably hold them responsible. I speculated that the view they continued to possess these capacities was a product of a widespread and intuitive, but false, view that the mind and its mechanisms is confined to the skull of agents. (Levy, 2012, p. 609).

Cova (2014) attempts to bypass Levy's objection by using what he calls the *direct Frankfurt counter-argument* (DFC). The DFC relies on the fact that Frankfurt cases such as *Voting* can directly elicit the intuition that their agent is morally responsible, so that comparison with a normal case is not needed. The argument runs this way:

The direct Frankfurt counter-argument:

1. The agent is morally responsible in the *Frankfurt* case.
2. However, due to the presence of the CI, the agent could not have done otherwise in the *Frankfurt* case.
3. Therefore, the PAP is false, since the agent in the *Frankfurt* case is morally responsible even though he could not have done otherwise.

In this argument, all references to the *Normal* case have disappeared and the contentious premise has been suppressed. As a result, it seems that the argument escapes Levy's criticism, since it no longer rests on the assumption that the presence of a CI makes no difference between the *Normal* and *Frankfurt* cases.

However, in an answer to Cova, Levy (2014) stresses that intuitions do not come from nowhere. Rather, they are likely to rely on implicit assumptions and inferences about the world, agents, and the human mind. Thus, it might be that intuitions about FSCs themselves rest on the same internalist prejudice, so that their reliability depends on the truth of this prejudice (if it is one). For this reason, appeal to direct intuitions about FSCs won't help, unless we show that internalism is true (a task that, we agree, might prove difficult) or that people's intuitions about FSCs are not based on a prior commitment to internalism.

2. Is there really an asymmetry between intuitions about FSCs and intuitions about FECs?

In this paper, we choose to take the second route and question the claim that people's intuitions about FECs are unreliable because they tacitly rely on some kind of internalism.

Casting doubt on this claim requires understanding Levy's argumentative strategy in its favor. We think that Levy's overall argument for this claim can be summed up in two steps:

- (i) Showing that intuitions about FSCs are inconsistent with other intuitions, and are as a result unreliable.
- (ii) Showing that what best explains these inconsistencies is the fact that people are tacitly committed to internalism.

Let's first begin by examining statement (i). Indeed, it is important to grasp what the supposed inconsistencies are before assessing whether (ii) is true. According to Levy, the main reason to think that our intuitions about FSCs are unreliable is that they are at odds with our intuitions about FECs, despite the fact that they both involve a counterfactual intervener. So, it would be more accurate to reformulate (i) in the following way:

- (i*) Showing that there are certain inconsistencies in our intuitions about cases involving counterfactual interveners (including both FSCs and FECs).

Levy's main reason to conclude that our intuitions about cases involving counterfactual interveners are inconsistent is that our intuitions about FSCs and FECs appear to contradict

each other. Indeed, for Levy, intuitions about FSCs are directly opposed to our intuitions about FECs (see for example Levy, 2011, p.177).

Note that mere difference in intuitions is not the same as a contradiction. We can have distinct intuitions about two cases without such intuitions being directly at odds with each other. Take for example the two following cases:

Compelled Voting: The initial situation is the same as in *Republican Voting*. However, this time, Connie does think deeply about the relevant policy. As a result of this thought, the neuroscientist intervenes, causing her to choose to vote Republican.

No Voting: This time, Connie has not yet made up her mind whether to vote Democratic or Republican. In fact, she is still at home, wondering if she should go vote or not. Unbeknownst to her, an evil but gifted neuroscientist monitors her neural states using a computer chip he has implanted in her brain. The chip gives the neuroscientist the power to intervene in Connie's neural processes. Connie is psychologically constituted so that it is a necessary (though not sufficient) condition of her deciding to go voting that she thinks deeply about the political importance of the vote; were Connie to think deeply about this policy, the neuroscientist would intervene, causing her to choose to stay at home but not go voting. But Connie does not think deeply about the vote's political importance and decides to stay at home on her own. The neuroscientist and his device play no role in bringing about Connie's decision to stay at home on her own.

In both cases, we have the intuition that Connie is not responsible for voting Republican. However, although these intuitions are different from our intuition in the *Republican Vote* case, there is no real conflict or opposition. That we judge Connie not responsible for voting Republican in the *Compelled Voting* and *No Voting* cases does not mean that our intuition that Connie is responsible for the same action in *Republican Voting* is misguided. Indeed, it is perfectly rational to have different intuitions about different cases as long as there are *relevant differences* between these cases.

Here, we focus on intuitions about moral responsibility. This means that there should be *morally relevant* differences between the cases. For example, the fact that we do not have the same intuition in the *Republican Voting* and *Compelled Voting* can be explained very simply by the fact that the neuroscientist is only a counterfactual intervener in *Republican Voting* while he actually intervenes in *Compelled Voting*. This means that the agent is able to exert guidance control upon his action in *Republican Voting*, but not in *Compelled Voting*.

The relevant difference is even clearer in the case of *No Voting*. What explains that Connie is responsible for voting Republican in *Republican Voting* but not in *No Voting*? Simply that Connie actually votes Republican in the first case, but not in the second case. As trivial as it sounds, we should not forget that one cannot be responsible for ϕ -ing if one did not actually ϕ . This will prove relevant later.

Moreover, the *No Voting* case also highlights an important condition for properly comparing intuitions about cases: that we ask the same question in all cases. For example, one could say that our intuitions about the *Republican Voting* and the *No Voting* are the same because we judge Connie responsible for voting Republican in the first case, and we judge her responsible for staying at home in the second. Thus, the alleged ‘difference’ between the two cases might only be an artifact due to the fact that we do not ask the right question. Given the structure of both cases, it might make more sense to compare intuitions about voting Republican in the first case to intuitions about staying at home in the second case. At least, these are the actions Connie actually performs while she could not have done otherwise.

With that in mind, we can clarify Levy’s main argument. His argument is that we have different intuitions about FSCs and FECs while we shouldn’t, given that *there is no morally relevant difference* between the two types of cases. All we have to do is understand how intuitions about FSCs and FECs are supposed to be different. In the rest of this section, we

consider, and assess, four ways Levy might characterize the difference between these intuitions.

1. One way one might characterize the difference is by claiming that (i) people agree that agents can gain capacities due to the mere presence of a CI in FECs, while (ii) they deny that agents can lose capacities due to the mere presence of a CI in FSCs.

We agree with (i). Jillian does seem to gain a capacity due to the mere presence of the CI in the *Phobia* case. As Levy puts it:

I argued that Jillian is responsible for failing to save the child. Jillian thinks that she can save the child. Moreover, I claimed, she is right: she has the capacity to save the child. But she has that capacity only due to the presence of the good-hearted neuroscientist. (Levy, 2012, p. 610).¹

However, we disagree with (ii). Indeed, it seems that, far from denying that the mere presence of a CI can change an agent's capacities, FSCs actually presuppose that the mere presence of a CI can change an agent's capacities. Indeed, is not the whole point of FSCs the idea that the mere presence of a CI is enough to *make the agent lose the capacity to do otherwise*? To take a more concrete example, is not the whole point of *Voting* the idea that the mere presence of the neuroscientist makes Connie lose the capacity to vote democrats? Clearly, people have the intuition that the mere presence of a CI in FSCs is enough to make agents lose capacities. Thus, this is not the most plausible reading of Levy's claim.

¹ Note that we agree with Levy that Jillian gains the capacity to save the child. One reviewer for this journal argues that she also gains the general capacity to swim. However, it is not clear: it is clearly stated in *Phobia* that Jillian is already a good swimmer, which means she already has the capacity to swim (for example, she can swim in shallow waters). If she had not already (and had no idea how to swim), the mere presence of the CI would not be enough to grant her the possibility to save the child.

2. A more plausible reading of Levy's claim can be formulated in terms of what Levy calls 'responsibility-underwriting capacities'. As he puts it:

proponents of FSCs accept that abilities can be lost due to the presence of a counterfactual intervener: after all, their cases turn on the claim that the ability to do otherwise is lost. It is for this reason that I have cast the discussion in terms of *responsibility-underwriting* capacities. The proponents of FSCs maintain, and I deny, that the relevant capacities are constituted entirely by intrinsic properties of the agent. My strategy is to show that responsibility-underwriting capacities can be gained through the mere presence of a counterfactual intervener, and that therefore we should think that they might be lost in the same kind of way. (Levy, 2012, p. 608.)

We just need to understand what Levy means by 'responsibility-underwriting capacities'. For example, is the capacity of saving the child in *Phobia* a responsibility-underwriting capacity? In one sense, yes, because Jillian could not be responsible for failing to save the child in *Phobia* if she did not have the capacity to save him. However, the capacity to save the child is not relevant when it comes to assessing Jillian's responsibility for, say, deciding to walk along the beach. We should then distinguish between capacities that are necessary for moral responsibility in certain circumstances (such as the capacity to swim, or the capacity to save the child) and capacities that are always necessary for moral responsibility (such as the capacity to respond to reasons). Let's call the former *circumstantial* responsibility-underwriting capacities and the latter *essential* responsibility-underwriting capacities.

One way to understand Levy's claim, therefore, is this in terms of *circumstantial* responsibility-underwriting capacities. Thus, the claim would be that (i) people agree that agents can gain *circumstantial* responsibility-underwriting capacities due to the mere presence of a CI in FECs, while (ii) they deny that agents can lose *circumstantial* responsibility-underwriting capacities due to the mere presence of a CI in FSCs.

However, statement (ii) is doubtful. The capacity supposed to be lost in FSCs is the capacity to do otherwise or, equivalently, the capacity ‘not to do X’, where X is what the agent actually does. But the capacity ‘not to do X’ is a necessary condition for the agent being morally responsible for actually failing to do X (for the capacity ‘not do X’ is essential for an agent to count as simply failing to X). However, it is not a necessary condition for the agent being morally responsible for doing something completely different, say doing Y. Thus, it a *circumstantial* responsibility-underwriting capacities. As a result, in all FSCs, there is at least one circumstantial responsibility-underwriting capacity that is lost. This means that Levy’s statement cannot be charitably interpreted as claiming that the advocate of FSCs is committed to denying that agents in FSCs cannot lose responsibility-underwriting (or even *circumstantial* responsibility-underwriting) capacities. For the claim to work, a narrower class of capacities should be singled out.

3. A more plausible reading of Levy’s claim is then a reading in terms of *essential* responsibility-underwriting capacities. Levy’s claim would be that (i) people agree that agents can gain *essential* responsibility-underwriting capacities due to the mere presence of a CI in FECs, while (ii) they deny that agents can lose *essential* responsibility-underwriting capacities due to the mere presence of a CI in FSCs.

This reading is suggested by Levy himself. He writes (2008, p. 224):

FSCs can show that agents are responsible for their actions, despite lacking alternative possibilities, only if we are justified in believing not only that the agents they feature act “on their own,” but also that their actions are an expression of *their* normal capacities for evaluating and responding to reasons. (Levy, 2008, p. 224.)

Here, Levy gives as an example of responsibility-underwriting capacities the capacity to evaluate and respond to reasons. Arguably, this is an *essential* responsibility-underwriting capacity. Levy's view, therefore, may be best characterized in terms of responsibility-underwriting capacities of this sort. Let's consider it.

Under this reading, statement (ii) is highly plausible. Let's compare *Republican Voting* with *Normal Voting*: in both cases, we judge Connie responsible for voting Republican. As a result, we must attribute to her all *essential* responsibility-underwriting capacities in both cases (if didn't, we wouldn't judge her responsible in the first place). And this means that we do not consider that the mere presence of the neuroscientist is enough to make her lose *essential* responsibility-underwriting abilities.

The problem, this time, is with statement (i). It is not so clear that we consider that the mere presence of the CI is sufficient to make Jillian gain *essential* responsibility-underwriting abilities. Granted, we have the intuition that, had the neurosurgeon not been present, Jillian would not have been responsible for failing to save the child. The mere presence of the neurosurgeon, therefore, must make Jillian gain something. However, it could be that this something is only a *circumstantial* responsibility-underwriting capacity, and not an *essential* responsibility-underwriting capacity.

In fact, there are reasons to think that we have the intuition that Jillian did not gain any *essential* responsibility-grounding capacities. Consider *Normal Phobia* – a case similar to *Phobia*, but in which there is no neurosurgeon. Now, consider a variation on this case in which, after deciding not to go save the child, Jillian continues her walk on the beach and assaults an old lady and steals her purse. Is Jillian responsible for assaulting the old lady and stealing the purse? The answer is 'yes': being pathologically afraid of deep water does not excuse you from such dark deeds. However, if Jillian in *Normal Phobia* can be responsible for such actions, then this means that all her *essential* responsibility-underwriting capacities

are intact: the only thing that prevents her from being responsible for failing to save the child are *circumstantial* responsibility-underwriting capacities, such as the capacity to swim. But if Jillian in *Normal Phobia* already has the required essential responsibility-underwriting capacities, then the presence of the neuroscientist in *Phobia* cannot make her gain an essential responsibility-grounding capacity: you cannot gain what you already have.

4. One last interpretation of Levy's claim is the following: (i) people agree that agents gain capacities due to the mere presence of a CI in FECs and lose capacities due to the mere presence of a CI in FSCs, but (ii) consider that the CI's presence has an effect on the agent's moral responsibility in FECs but not in FSCs. In other words, people agree that the mere presence of a CI can have an effect on agents' capacities (in both FECs and FSCs) but consider this fact as relevant for agents' moral responsibility only in the case of FECs. Thus, people seem to treat both kinds of case asymmetrically.

This argument supposes that we are comparing pairs of cases. Let's first take the *Normal Phobia / Phobia* pair. People agree that adding a CI in *Phobia* makes the agent gain new capacities. As a result, they accept that our attributions of moral responsibility should differ between the *Normal Phobia* and *Phobia* cases (they consider that the agent is not responsible in *Normal Phobia*, but is responsible in *Phobia*).

Consider now the *Normal Voting / Republican Voting* pair. In this case, people agree that adding a CI in *Republican Voting* makes the agent lose certain capacities. However, they do not seem to accept that our attributions of moral responsibility should differ between the *Normal Voting* and *Republican Voting* cases (they consider that the agent is responsible in both cases).

Comparing our reactions to the two pairs of case, Levy could propose the following argument: we treat modification in the agent's capacities due to the mere presence of a CI as

relevant in the case of FECs, but not in the case FSCs. But there is no morally relevant difference between the two kinds of cases. As a result, our intuitions about cases involving CI are unreliable.

However, we saw earlier the importance of comparing cases that are structurally identical and do not involve obvious morally relevant differences when trying to reach such conclusions. However, there seems to be morally relevant differences between the two pairs of cases: while in *Normal Voting* and *Republican Voting*, the agent does exactly the same thing (voting Republican), she doesn't do the same thing in *Normal Phobia* and *Phobia*. In *Phobia*, Jillian *fails to save* the child., In *Normal Phobia*, by contrast, she does not, strictly speaking, fail to save the child, or at least not in the same way. Ordinarily, failing to ψ seems to imply that one could have ψ -ed. Saying that Harry is constantly failing to go to the moon is certainly not something we would say unless we thought that it was possible for Harry to go to the moon (if we thought, for instance, that Harry was an astronaut). The same thought can be applied to *Normal Phobia*; given that it was impossible for Jillian to save the child in this situation, we cannot claim that she actually failed to save the child. And so given that Jillian does not do the same thing in the two cases, we may have a basis for claiming that there is a morally relevant difference between them. All this seems to suggest that the four cases that Levy is comparing are not morally equivalent, and that no conclusion can be drawn from their comparison.

Another problem can be raised against the question asked in relation to each pair of cases: in the *Phobia* cases, the agent gains the ability to 'save the child', and Levy consider people's judgments about Jillian's responsibility for 'failing to save the child'. However, in the *Voting* cases, the agent loses the capacity to 'vote Democrats', but Levy focuses on whether Connie is responsible for 'voting Republicans'. These are two very different questions. For both questions to be equivalent, we should rather ask whether Connie is

responsible for ‘failing to vote Democrats’. This makes a difference, since we have reasons to think that Connie is responsible for failing to vote Democrats in *Normal Voting*, but not in *Republican Voting* (indeed, since she could not have voted Democrats in *Republican Voting*, she does not even qualify as having failed to vote Democrats).²

A good comparison would rather involve comparing *Phobia* and *Normal Phobia* to the following cases:

Bad Samaritan: A case in which Jillian does not suffer from a phobia (she’s a normal agent), and in which there is no CI. Still, Jillian decides not to save the child.

Forbidden Samaritan: A case similar to *Bad Samaritan*, but in which a CI would have prevented Jillian from deciding to save the child, had she shown any inclination to do so.

This is the FSC-pair that is truly symmetrical to Levy’s FEC-pair. However, in this case, we have (i) the intuition that the mere presence of the CI is enough to make Jillian lose a capacity (the capacity to save the child), and (ii) the intuition that Jillian is responsible for failing to save the child in *Bad Samaritan* but not in *Forbidden Samaritan*. Thus, it seems our intuitions about FSCs are perfectly symmetrical to our intuitions about FECs.

We can conclude that it does not seem that our intuitions about FECs and FSCs conflict, once we have eliminated all morally relevant differences between the two, and asked the right questions. This excludes one reason to think that our intuitions about cases involving CI are unreliable.

However, there is still one argument open to Levy. During all this discussion, we have agreed that it was obvious that the mere presence of a CI was enough to make agents gain or

² Similarly, if one insists that the relevant question in *Republican Voting* is “is Connie responsible for Voting Republicans”?, then one should consider that the relevant questions in *Phobia* is not “is Jillian responsible for

lose certain capacities. Levy could then ask: if we agree on that, how can we rule out the possibility that the mere presence of a CI is enough to make agents lose essential responsibility-underwriting capacities, such as the capacity to respond to reasons? Why treat these capacities differently? Is it not evidence that our intuitions about essential responsibility-underwriting capacities (and, thereby, about FSCs) rely on controversial internalist commitments?

3. Do intuitions about FSCs rest on an internalist bias?

The answer to this question is: not necessarily. There are ways to justify this difference in treatment that do not obviously rely on any sort of internalist thesis about the mind.

To understand why, let's begin with a distinction between two kinds of capacities. Writing about 'abilities' rather than 'capacities', Maier (2010) makes a distinction between *general* and *specific* abilities:

The distinction between general and specific abilities may be brought out by way of example. Consider a well-trained tennis player equipped with ball and racquet, standing at the service line. There is, as it were, nothing standing between him and a serve: every prerequisite for his serving has been met. Such an agent is *in a position to serve*, or has serving as an *option*. Let us say that such an agent has the *specific ability* to serve.

In contrast, consider an otherwise similar tennis player who lacks a racquet and ball, and is miles away from a tennis court. There is clearly a good sense in which such an agent has the ability to hit a serve: he has been trained to do so, and has done so many times in the past. Yet such an agent lacks the *specific ability* to serve, as that term was just defined. Let us say that such an agent has the *general ability* to serve. (Maier, 2010).

failing to save the child?", but "is Jillian responsible for continuing to walk on the beach?". Here again, both

Following Maier, we can easily draw a similar distinction between *general capacities* and *specific capacities*: while the general capacity to X might survive the absence of possibility to X (in a given situation), the specific capacity to X presupposes the possibility to X (in a given situation). Put differently, having the specific capacity to X at t requires that there be at least a possible world in which the agent does X that is accessible to the agent at t, while the general capacity to X might survive the absence of such a possible world.

This relative insensitivity of general capacities to which possibilities are open to the agent might be explained in different ways. According to a strong realist view, to which we are sympathetic (see below), general capacities are dispositions, and dispositions exist independently from their actual and possible manifestations. If we adopt such a view, then general capacities exist independently of what does or could happen: actual or potential manifestation of these general capacities are *explained* by these general capacities but are not required for such general capacities to exist.

On a weaker view, general capacities might be reduced to the occurrence of certain events in a fairly unrestricted set of situations, such that an agent A has the capacity to X just in case A would perform X if she tried to X in (e.g.) normal circumstances. On this view, the presence of an unmanifested general capacity at a world *w* depends on what happens at various other worlds. The clause ‘in normal circumstances’ is meant to allow them to survive even when no immediate possibility to do X is open. For example, even if there is no water nearby, one might still be said to have the general capacity to swim given that one would swim if one tried to and there was water nearby. The presence of a specific capacity, by contrast, would depend, on this weaker view, on the occurrence of certain events in a *much more* restricted set of situations, such that an agent A has the capacity to X *in situation S* just in case she would perform X if she tried to X *in S*. In the water case, although one might have

questions yield the same answer: “yes”.

the general capacity to swim, one would fail to have the specific capacity to swim in a situation in which no water is present.

At least for now, whatever view of capacities one ends up choosing, it is clear that general capacities can survive the absence of certain possibilities, especially when these possibilities are closed by ‘unusual circumstances’, such as the ones encountered in FECs and FSCs. By contrast, because specific capacities depend on the existence of such possibilities, they will be highly sensitive to changes in these possibilities. Given that CI (by definition) only exert direct influence on what happens at other worlds, and that their influence on agent’s actual capacities only depends on their influence at these worlds, it is no wonder that agents’ specific capacities can be dramatically modified by the mere presence of CIs. However, for the same reasons, we should not expect CIs to have much effect on an agent’s general capacities.

Let’s now get back to our distinction between essential responsibility-underwriting capacities and circumstantial responsibility-grounding capacities. It seems to us that most essential responsibility-underwriting capacities, such as the capacity to respond to reasons, are general capacities. If so, then, we should expect essential responsibility-underwriting capacities to be highly indifferent to the presence of a CI. This might explain why we have the intuition that the presence of a CI cannot alter an agent’s essential responsibility-underwriting capacities: it is because we think that the existence of such capacities does not depend on the occurrence of their manifestations in a restricted set of situations.

However, Levy might be tempted to reject the distinction we drew between general and specific capacities by arguing that this distinction itself presupposes an internalist conception of the mind, or at least of agents and their capacities. For example, discussing a related (though different) objection by Clarke (2011) that rests on the distinction between

intrinsic and extrinsic capacities, Levy argues that this very distinction “is covertly circular” because, in following this distinction, we “individuate mechanisms by reference to the very intuitions that the mechanisms are then supposed to underwrite” (Levy, 2012, p. 618). In other words, the distinction between intrinsic and extrinsic capacities could be based on the same internalist prejudice Levy perceived in FSCs, and this might also be true for the distinction between general and specific capacities.

Such an answer strikes us as implausible, though, because the distinction between general and specific is not specific to our conception of the mind, or to our conception of agents and their capacities. Indeed, this distinction can be grounded in a more fundamental distinction at the level of *dispositions*.

Dispositions, it used to be said, are merely a matter of what would happen if certain conditions were met. A fragile glass, the story goes, is disposed to break just in case, if it were struck, it would break. The so-called ‘conditional analyses’ resulted from the observation that dispositions appear closely related to conditionals. Simple conditional analyses of the sort just given for fragility, however, are widely rejected on the basis of at least three sorts of counterexamples. First, there are cases – the so-called ‘finkish’ cases – where an object has a disposition at a given time but would *lose* its disposition in its stimulus conditions (Martin, 1994). Using an example that Levy himself gives (2008, p. 227), a vase can possess the property of being fragile right now, while sitting on its shelf, and this *regardless* of the fact that if it was struck it wouldn’t break due the intervention of some external factor. The second sort of case – a ‘masking’ case – is a case where an object has a given disposition, would keep its disposition in its stimulus conditions, but would fail to manifest it in the relevant conditions. A fragile glass, for instance, can be wrapped up in such a way that it would not break if struck; and although it would fail to break in this situation, it is quite clear that it doesn’t thereby lose its fragility (Johnston, 1982). The last sort of case – a ‘mimicking’ case –

is where an object, although lacking a given disposition, would behave as if it had it in a situation relevant to the kind of disposition in question. Let's suppose that being shy is the disposition not to talk much at social gatherings and suppose that Bill, a not very shy person, is paid to act shyly at all the parties that he will attend in the future. In such a case, it is the case that, if Bill were to attend a party, he would not talk much at it, but it is clear that he would not thereby count as shy.

A common response to these cases is to move from an analysis of dispositions in terms of conditionals to an analysis of dispositions in terms of *actual properties* that objects can possess at a time/world regardless of what happens at other times/worlds. It is now standard to say that dispositions can remain unmanifested even in conditions in which they are supposed to manifest themselves, suggesting that conditional statements should be taken with a pinch of salt, as mere heuristics perhaps, when it comes to understanding dispositions (see, e.g., Molnar, 2003). Quite uncommon, however, is to remark that the conditional analysis seems to be true of a certain *use* of the expression 'is disposed to' (Naar, 2013, p. 84). What we have in mind here are primarily cases of disposition ascriptions made by committed Humeans, such as Gilbert Ryle, who do not think that there are dispositional properties at all, over and above the relevant conditionals, and who use dispositional predicates in, as it were, a purely behaviouristic fashion. We suspect, however, that such a use can be found even in contexts where no commitment to the simple conditional analysis or anything like it is implicitly accepted by the relevant parties. Taking a case where a given disposition is said to be mimicked, it still makes sense, it seems, to say that the object is *disposed* to (or, perhaps better, *tends* to) act in a certain way if subjected to a given stimulus. Similarly, it seems to make sense to say that a glass that is wrapped up in some packing material is *not* disposed to break when struck. It seems to us, therefore, that there is a use of 'disposed to' that doesn't commit the speaker to the presence of a unique property explaining the events that he

describes. Let's call dispositions of the realist, property sort 'powers' and dispositions of the chiefly antirealist sort just talked about 'Humean dispositions'.

There are various possible combinations of powers and Humean dispositions. An object can have the power to break, say, without having the Humean disposition to break. And an object can have the Humean disposition to break without having the power to break. In many cases, however, objects have both the power to break and the Humean disposition to break (e.g., a fragile glass that is not wrapped up). Now, it is always possible for one to say that, in the absence of evidence for the presence of a given power, an object can nonetheless *gain* the corresponding Humean disposition if it would behave in a certain way in a given situation. The acquisition of the Humean disposition, moreover, can be due to the presence of an external factor that would, in that situation, 'make sure' that the object behaves in the relevant way, and this regardless of whether it possesses the corresponding power. Whether or not a given object is indeed fragile, it can gain a Humean disposition by virtue of the fact that it would break if struck in a certain situation (due to the presence of sorcerer, for instance).

Notice now that, although the presence of counterfactual factors typically does not have an impact on the possession of powers, it always has an impact on the possession of Humean dispositions. A committed Humean, of course, might reply that making powers insensitive in this way to counterfactual factors commits us to an implausible 'actualism' about dispositions. The problem is, however, that actualism about dispositions – at least those that are plausibly construed as powers – has a lot going for it (e.g., Mumford, 1998, Molnar, 2003).

It is now possible for us to reformulate our objection to Levy's argument in more metaphysically fundamental terms. It seems to us that, although circumstantial responsibility-grounding capacities can be either powers (such as the capacity to swim in general) or

Humean dispositions (such as the capacity to swim at a particular moment *m* and to save the child in Jillian's particular situation), most (if not all) essential responsibility-underwriting capacities are powers (such as the capacity to respond to reasons). But, if they are, then, as powers, they should not be very sensitive to the presence of a CI. Granted, agents in FECs such as *Phobia* do gain capacities due to the mere presence of a CI, but these capacities are only Humean dispositions, so that no lesson about powers can be drawn from them.³

Consequently, we believe that what Levy needs in order for his case to work is not a rejection of internalism. What he needs, we contend, is the denial of the actualism about dispositions that we have motivated; he must accept the view that there are no genuine powers, only Humean dispositions. Given the plausibility of actualism about powers – which Levy himself seems sympathetic to (Levy, 2008, pp. 227-228) – and the plausible claim that agential capacities are a species of dispositions, it is quite reasonable to conclude that not all capacities are Humean, i.e. that some capacities are powers. Unless Levy gives an argument to the effect that capacities are special in that none of them are powers (that they all are Humean), or an argument to the effect that capacities are not a species of dispositions, Levy fails to have the dialectical advantage he claims he has.

To conclude, the distinction between powers and Humean dispositions, which arguably grounds the corresponding distinction between general and specific capacities, gives us reasons to resist the conclusion that considering that certain capacities can be affected by the mere presence of a CI but that others cannot is absurd or arbitrary. Because there are two different species of capacities (and dispositions), it makes perfect sense not to treat all capacities alike.

³ Additionally, the distinction between powers and Humean dispositions seems to allow us to explain the disagreement between Levy and some of his critics: while Levy use 'abilities' in a sense that includes Humean

4. Conclusion

We have now reached the main goal of this paper: showing that intuitions about FSCs are neither unreliable nor grounded in a tacit commitment to internalism. However, what does this conclusion imply for the relevance of FECs in the debate over PAP? Does it imply that FECs are unable to cast doubt on intuitions about FSCs? Let's close this paper by examining the matter.

One reviewer for this journal raised the following objection: most CI in FECs and FSCs are only, as it were, 'one night stands'. But what if we imagined a FEC in which the CI accompanies the agent all her life (or, at least, a good chunk of it)? Wouldn't that change our intuitions about the impact of CI in agent's powers?

Probably. Consider the following case. Mary does not know how to swim. In fact, she is unable to. However, one day, for a short moment, her fairy godmother gives her a magic potion. If Mary drinks the potion and decides to go swimming, she would suddenly know how to do swim, and would swim like a champion. But the potion's effect will only last one hour, and the potion must be drunk fresh, and will only work the day it has been prepared. Despite this kind gift, Mary decides not to go swim and the potion hits its expiration date. Did Mary have the power to swim during this time? Not really; she merely has the possibility to swim.⁴

But what if, rather a potion, her godmother gave her a necklace that has the same effect: it activates whenever Mary decides to swim and grants her the possibility to swim like a champion as long as she is in the water. Let's now suppose that Mary keeps the necklace all

dispositions, his critics only use this term to refer to powers. See: Haji and McKenna M (2011) and Cova (2014).
⁴ Could it be said that Mary has at least the Humean disposition to swim? If we consider that Mary has the Humean disposition to swim only if there is an accessible possible world in which she swims, then it does not seem to be the case. Still, she has access to the possible worlds in which she has drunk the potion. In these worlds, she has access to worlds in which she swims. So, our guess would be that Mary has the Humean

the time, and wherever she goes, even if she never decides to go swimming (she really hates water). Wouldn't we be more inclined to say that she actually has the power to swim (that she 'can swim' in the general sense of the expression)? Probably.⁵

Suppose that this is the case. What would it imply regarding the debate over PAP? Could this sort of FEC undermine FSCs? Not by showing that our intuitions about such cases rely on controversial internalist commitments. Indeed, this would in fact show the direct opposite, by showing that we are ready to accept that factors external to the agents can affect its general capacities and powers.

Could such a case be used to cast doubt on the distinction between general and specific capacities? We do not think so. As we saw earlier, some authors distinguish between the two kinds of capacities by claiming that A has the general capacity to X just in case A would perform X if she tried to X in a fairly wide range of circumstances, while she has the specific capacity to X at it just in case she would perform X if she tried in the situation she is in. This distinction could readily account for the difference between the potion and necklace cases: while the potion is not part of the usual circumstances of Mary's life, and so cannot ground a general capacity to swim, the necklace is. So, acknowledging that long-term CIs can modify an agent's capacities does not threaten the distinction we drew between two kinds of capacities (and two kinds of dispositions).

One possible move, then, would be to argue that, because we accept that the mere presence of a CI is enough to modify an agent's general capacities in such cases, then we should also accept that the mere presence of a CI is enough to modify an agent's general capacities in FSCs. However, once again, we must make sure that we compare cases that are

disposition to have the Humean disposition to swim. However, we can easily grant that Mary has the Humean disposition to swim: after all, nothing in the present argument depends on this claim.

⁵ After all, we tend to say that Green Lantern has the power to fly, even if this power depends on his having a 'magic' ring, the power of which also depends on the existence of a Lantern, the power of which depends on the existence of an even bigger Lantern. But note that we would not say that Green Lantern's girlfriend has the power to fly only because she has often the possibility to steal Green Lantern's ring and use it to her purpose.

structurally similar. We should therefore say that the mere presence of a CI is enough to modify an agent's general capacities in FSCs *in which the CI accompanies the agent all her life (or a good chunk of it)*. But this has no implication whatsoever for standard FSCs, in which CIs only intervenes punctually.

Another solution, also suggested by a reviewer, might simply be for Levy to argue that, since we have accepted that the agent in *Phobia* is responsible in virtue of gaining a capacity, we should also accept that agents may be excused in virtue of losing a capacity of whatever type is lost in FSCs such as *Republican Voting*.

We already addressed this issue in section 2. However, it might be useful to clarify our answer. First, as we pointed out, the lesson we drew from *Phobia* cannot be extended to any FSCs: they can only be extended to FSCs that are structurally similar to *Phobia*, that is, cases in which the capacity that is lost actually determines whether or not the agent performs the relevant action. *Republican Voting* is not such a case, since losing the ability to vote otherwise does not prevent one from voting Republican. Second (and this might be only a different way of making the same point), we know why gaining a capacity makes the agent suddenly responsible for failing to save the child in *Phobia*: it is because, in the absence of this capacity, it would not even be true that the agent would have failed to save the child. So, the only lesson we can draw from *Phobia* is that capacities that are gained (or lost) due to the mere presence of a CI have an impact on an agent's responsibility for ϕ -ing only to the extent that they are necessary for the agent to count as ϕ -ing. However, in FSCs such as *Republican Voting*, the capacity that is lost does not change anything to what the agent does: whether or not she is able to vote Democrat does not change the fact that Connie votes Republican.

As a result, it is not clear that FECs can be used to refute the claim that agents in FSCs are morally responsible for their actions. This does not mean that it would be impossible for someone to eventually come up with a clever way to use them for this purpose.⁶ However, we think that, even if FECs ultimately turn out to constitute a good objection to FSCs, it won't be in virtue of showing that our intuitions about cases involving counterfactual interveners are either unreliable or grounded in some controversial internalism about the location of mental states. Rather, our intuitions in such cases are based on a basic distinction between two kinds of dispositions. Attacking intuitions about FSCs on this base would then require one to dabble in the metaphysics of dispositions.

REFERENCES

- Clarke, R. (2011). 'Responsibility, Mechanisms, and Capacities', *The Modern Schoolman* 88, pp. 161-169.
- Cova, F. (2014). 'Frankfurt-Style Cases User Manual: Why Frankfurt-Style Enabling Cases Do Not Necessitate Tech Support', *Ethical Theory and Moral Practice* 17, pp. 505-521.
- Fischer, J. M. (1994). *The Metaphysics of Free Will*. Oxford: Blackwell Publishers.
- Frankfurt, H. (1969). 'Alternate Possibilities and Moral Responsibility', *Journal of Philosophy* 66, pp. 829-839.
- Ginet, C. (1996). 'In Defense of the Principle of Alternate Possibilities: Why I Don't Find Frankfurt's Argument Convincing', *Philosophical Perspectives* 10, pp. 403-417.

⁶ However, it should be noted that FECs are not only used to defeat FSCs. Indeed, Levy (2008) also rely on FECs to argue that external circumstances can make a difference to agents' general capacities (and, more particularly, to agents' essential responsibility-underwriting capacities), so that the possession of such capacities

- Haji, I. and McKenna, M. (2011). 'Disabling Levy's Frankfurt-style Enabling Cases', *Pacific Philosophical Quarterly* 92, pp. 400-414
- Johnston, M. (1992). 'How to Speak of the Colors', *Philosophical Studies* 68, pp. 221-263
- Levy, N. (2008). 'Counterfactual Intervention and Agents' Capacities', *Journal of Philosophy* 105, pp. 223-239.
- Levy, N. (2011). *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. Oxford University Press.
- Levy, N. (2012). 'Capacities and Counterfactuals: A Reply to Haji and McKenna', *Dialectica* 66, p. 609.
- Levy, N. (2014). 'Countering Cova: Frankfurt-style cases are still broken', *Ethical Theory and Moral Practice* 17, pp. 523-527.
- Maier, J. (2010). 'Abilities', in E. Zalta (ed), *Stanford Encyclopedia of Philosophy*.
- Martin, C. B. (1994). 'Dispositions and Conditionals', *Philosophical Quarterly* 44, pp. 1-8
- Molnar, G. (2003). *Powers*. Oxford University Press.
- Mumford, S. (1998). *Dispositions*. Oxford University Press.
- Naar, H. (2013). *A Defence of Sentiments: Emotions, Dispositions, and Character*. Doctoral Thesis, University of Manchester.
- Pereboom, D. (2001). *Living without Free Will*. New York: Cambridge University Press.
- van Inwagen, P. (1983). *An Essay on Free Will*. Oxford: Clarendon Press.

is a matter of luck. As shown in our discussion of long-term CIs, we are happy to grant this point. Our focus in this paper is only Levy's criticism of intuitions about FSCs.