The Future of the Causal Quest

Provisionally forthcoming in *Blackwell Companion to Free Will*

1. Introduction

In one sense, questions about the nature of free will can be understood as ethical or normative questions. We arguably care about being free to the extent that we care about being morally responsible. In this way, the nature of free will is inextricably tied to the nature of moral responsibility. But questions about the nature of free will can also be understood as metaphysical questions. Though many take free will to be the kind of freedom or control required to be morally responsible, it's understood as a necessary metaphysical condition for moral responsibility.[1] Given that free will is a metaphysical concept, it stands to reason that work on other metaphysical concepts can shed light on the nature of free will. And work on the metaphysics of laws, dispositions, and abilities has certainly impacted and informed our understanding of free will in a meaningful way. However, there has been a relative reluctance to turn to the literature on the metaphysics of causation. This reluctance is perhaps best captured by Peter van Inwagen, who famously wrote: "Causation is a morass in which I for one refuse to set foot. Or not unless I am pushed" (van Inwagen 1983: 65).[2] But this appears to be changing, for several philosophers have recently begun to wade into the morass.[3] And as we'll see, though it's becoming increasingly agreed upon that causation and free will are related, it's much less clear how we should go about incorporating the metaphysics of causation into our accounts of free will. In this chapter, I will look at three recent attempts to draw lessons about free will from the causation literature: Oisín Deery and Eddy Nahmias' (2017) account of interventionist causation and manipulation arguments, Carolina Sartorio's (2016) actual causal sequence account of free will, and Sara Bernstein's (forthcoming) analysis of the relationship between causal proportionality and moral responsibility. What follows is far from an exhaustive analysis of the current work on causation and free will and in focusing on these particular views I've ignored many others. What I find compelling about these particular views is that though they represent three very different ways of incorporating work on causation into discussions of free will, they all face real challenges about how best to conceive of the relationship between the metaphysical and ethical questions regarding the nature of free will. And by reflecting on the different ways those working on free will can utilize the research on causation, and on the questions about the interplay between metaphysics and ethics that these approaches raise, we can reveal new and interesting avenues for future research not only on the relationship between causation and free will but on the metaphysics of free will more generally.

2. Deery and Nahmias' interventionist response to manipulation arguments

In their recent paper, "Defeating Manipulation Arguments: Interventionist Causation and Compatibilist Sourcehood," Oisín Deery and Eddy Nahmias (2017) use an interventionist account of causation to defend compatibilists against one of the most serious objections on the market:

[1] Often contrasted with an epistemic condition, which refers to the kind of knowledge or awareness required to be morally responsible.

[2] Those who defend agent-causal libertarian views (e.g. O'Connor 1995) and event-causal libertarian views (e.g. Kane 1996) of free will are notable exceptions.

[3] For example, Dana Nelkin (2011) has developed a compatibilist account of agent causation and Kadri Vihvelin (2013) relies on the metaphysics of causation in defending her dispositionalist account of free will.

manipulation arguments. There are many different kinds of manipulation arguments, but the most familiar instances share a common structure, which Michael McKenna has helpfully laid out:

> 1. If S is manipulated in manner X to A, then S does not A of her own free will and is therefore not morally responsible for A'ing.
> 2. An agent manipulated in manner X to A is no different in any relevant respect from any normally functioning agent determined to do A from (CAS) [the compatibilist-friendly agential structure].
> 3. Therefore, if S is a normally functioning agent determined to A from CAS, she does not A on her own free will and therefore is not morally responsible for A'ing (McKenna 2008, 143).

Deery and Nahmias use a case drawn from Alfred Mele's manipulation argument (2013) to develop their response.

> First, imagine Danny. One evening in 1986, Danny's parents made love, hoping to conceive a child. They got lucky. A zygote was formed (at time t_1), and nine months later Danny was born. Thirty years later, Danny is walking down a deserted street and he finds a wallet with the owner's ID in it and $500. Danny takes himself to have good reasons for keeping the money, but also for returning the wallet. He deliberates for a while, and in the end he decides to keep the money, and he does so (at time t_{30}). Assume that this occurs in a *deterministic* universe—that is, a universe in which, for each event E, the laws of nature and some set of events that occurred prior to E are such that these events cause E to occur with probability 1. If determinism is true, then some set of events prior to Danny's act of stealing the wallet at t_{30} are (together with the laws) such that they cause his deliberating and acting in that way, at that time, with probability 1. (Deery & Nahmias 2017: 1257)

Compare this to a different case:

> …a powerful Goddess, Diana, has the power to know what will happen in the future and to act in ways that ensure that specific events occur in the distant future. Diana has these abilities in part because she exists in a deterministic universe and is able to get enough information about events occurring in it (e.g., at t_1) to deduce exactly what she needs to do at that time to ensure that a particular event occurs thirty years later. In this case, Diana assembles atoms in a specific way at t_1 so as to create a zygote that develops into a child, grows up, finds a wallet thirty years later, and at t_{30} decides to keep the money it contains. For some reason, Diana wants to ensure that this event occurs at t_{30}, and she possesses the power to alter events at t_1 precisely so that she ensures that it does occur.
>
> As it turns out, the life of this intentionally created person (whom we will call Manny since he is *Man*ipulated) follows the exact same course as the life of deterministic Danny (as described above). Manny is no different from Danny when it comes to his abilities to consider options, weigh reasons, and make decisions about whether to steal the money. (Deery & Nahmias 2017: 1257).

The manipulation argument can proceed as follows. Intuitively, Manny does not have free will and is not morally responsible for stealing the money, and more generally, any agent who is manipulated in the way that Manny is does not have free will. Deery and Nahmias call this the NoFW Premise. But there is no principled difference between Manny and Danny—they meet all the same compatibilist sufficient conditions for free will—and there is no difference between the kind of manipulation featured in the case of Manny and the truth of determinism when it comes to free will and moral responsibility.[4] Deery and Nahmias call this the NoDif premise. So, Danny doesn't have free will and isn't morally responsible for stealing the money, just as no agent has free will and is morally responsible if determinism is true.

There are many different avenues a compatibilist can take in responding to manipulation arguments. McKenna calls the strategy of denying the NoFW Premise taking the "hard-line" while those who deny the NoDif premise take the "soft-line" (McKenna 2008). McKenna favors taking the hard-line, since it appears that soft-line responses are mere stop-gap solutions that only keep the incompatibilist at bay temporarily. A compatibilist can point to a condition for free will and moral responsibility that the manipulated agent does not meet that the determined agent does, but it seems in principle possible for the defender of the manipulation argument to augment the cases such that the manipulated agent meets the condition in question as well. But while taking the soft-line can (arguably) only succeed temporarily, taking the hard-line is no easy path; it requires the compatibilist to accept the (arguably) counterintuitive claim that Manny and those manipulated like him do in fact have free will and are morally responsible for the actions for which they were manipulated to perform.

Deery and Nahmias argue that it's possible for the soft-line to be more than a mere stop-gap defense in the face of manipulation arguments. In their recent essay, they reject the NoDif Premise and argue that there is an in principle difference between Manny (and those who are manipulated like him) and Danny (and those who exist in a world where determinism is true). They also contend that there is no way (or easy way) for the defender of manipulation arguments to alter their case to sure up the NoDif Premise in response to their soft-line defense of compatibilism. Deery and Nahmias rely on interventionist approaches to causation to reject the NoDif Premise.[5] Interventionism has a growing influence in the causation literature, and has figured in several recent discussions of agency and free will.[6] Deery and Nahmias argue that it can also be used to develop an account of the minimally sufficient conditions for free will that are immune to manipulation arguments.

Interventionism about causation involves constructing models that represent counterfactual relationships among events. We first assign variables to the event-types in question, and we can set these variables to different values in order to represent particular event-tokens (Woodward 2003). If we want to know whether the output of Danny (or Manny's) CAS caused him to steal the money, we first assign variables to the output of his CAS (X) and the stealing of the money (Y). Next, we can perform an intervention to determine whether X caused Y, which involves setting X to a different value and seeing whether such an intervention reliably changes the value of Y. For example, we can imagine that the output of Danny's CAS was the decision not to steal the money. In this case, it seems likely that this intervention would reliably result in a change in the value of Y, i.e. the stealing of the money wouldn't occur. We can also make finer-grained interventions, where we imagine that the output of Danny's CAS is the decision to steal the money but takes on a slightly weaker motivational force. Given this intervention, the value of Y will likely be different as well—perhaps Danny will delay in stealing the money so the event of stealing the money takes place at a slightly different time.

---

[4] See Frankfurt (1971), Wolf (1990), and Fischer and Ravizza (1998) for examples of different accounts of the minimally sufficient compatibilist conditions for free will.

[5] For examples interventionist accounts of causation, see Pearl (2009) and Woodward (2003).

[6] Ismael (2013, 2016), Roskies (2012), and Campell (2010).

On an interventionist approach to causation, the output of Danny (and Manny's) CAS counts as a direct cause of Danny (and Manny) stealing the money. This is because an intervention on X would reliably change the value of Y. However, being a direct cause of an event is distinct from being the *causal source* of the event, and according to many, causal sourcehood is what we're after when we seek to determine moral responsibility, praise, and blame (Cartwright 1979). According to Deery and Nahmias, for a cause to be the causal source of a given event, that cause must bear the strongest causal invariance relation to it among all the events that are causally relevant (Deery & Nahmias 2017). Deery and Nahmias characterize causal invariance as follows:

> A causal invariance relation, R_1, that obtains between two causal variables, X and Y, is stronger than another such relation, R_2, obtaining between Y and another of its prior causal variables—for instance, W—iff:
>
> (1) holding fixed the relevant background conditions, C, R_1 predicts the value of Y under a wider range of interventions on X than R_2 does under interventions on W; and
> (2) R_1 predicts the value of Y across a wider range of relevant changes to the values of C than R_2 does. (Deery & Nahmias 2017: 1262-1263)

Deery and Nahmias argue that in order for an agent to be free and thus morally responsible for a given action, the output of an agent's CAS must be the causal source of that action, in that it must bear the strongest causal invariance relation to that action. In the case of Danny, his CAS plausibly is the causal source of his decision to steal the money. We've already seen how changes in the value of the variable representing the output of Danny's CAS result in a change to the value of the variable representing Danny's stealing the money. And because Danny is an intentional agent, his decision to steal the wallet would lead to him to do so across many changes in background conditions. Danny would likely steal the money if it was in a slightly different location, if it were raining instead of snowing, if the wallet was black instead of brown, etc. Of course, there will be some changes in the background conditions that would result in Danny not stealing the money, i.e. if a police officer was standing nearby.

This is all true of Manny as well. In the case of both Danny and Manny, the relationship between the variables representing the outputs of their CAS and the variables representing their stealing the money is equally strong. But there is a stronger causal invariance relation between Diana's meddling and Manny's stealing, according to Deery and Nahmias.

> The relation between Diana's decision and Manny's action is such that, across a maximally wide range of changes to the background conditions, the variable representing Manny's stealing does not change in value without a change in the value of the variable representing Diana's decision, and changes in the value of the variable representing Diana's decision correspondingly change the value of the variable representing Manny's decision to steal... (Deery & Nahmias 2017: 1264)

Importantly, this is because Diana intends for Manny to steal the money at t_{30} and she is able to ensure that he does so. If Diana is able to ensure that Manny steals the money at t_{30}, then there's no change in the background circumstances that could alter the value of the variable representing Manny stealing the money. When evaluating the relationship between the output of Manny's CAS and his stealing the money, we must ignore Diana and all other causally relevant events—we are only concerned with the output of Manny's CAS and his stealing the money. And just as there are changes to the background circumstances that would stop Danny from stealing the money, i.e. if a police officer was standing

nearby, these changes would stop Manny as well. Thus, Diana's decision is the source of Manny stealing the wallet, not the output of his CAS.

From here, Deery and Nahmias are able to develop a soft-line response to the manipulation argument. They have isolated a relevant difference between Danny and Manny—while the output of Danny's CAS is the causal source of his stealing the money, the output of Manny's CAS is not. Thus, they can reject the NoDif premise both as it pertains to Danny and Manny and the more general claim that there is no difference between the kind of manipulation featured in Manny's case and the truth of determinism when it comes to free will. In this way, Deery and Nahmias have put forth a soft-line strategy that is meant to be much more than a stop-gap response to manipulation arguments, for they have isolated an in principle difference between manipulation and determinism,[7] while the former renders agents unable to be the causal sources of their actions, the latter does no such thing.

There is much to say about Deery and Nahmias' innovative approach to manipulation arguments. First, their use of an interventionist approach to causation is emblematic of a current trend in the work on free will. John Campbell (2010) uses interventionism to explain a notion of mental causation, Adina Roskies (2012) uses it to construct an account of self-authorship, and Jenann Ismael (2013, 2016) relies on it to defuse the threat of determinism to free will. As work on interventionist approaches to causation continues to advance, the extent to which these views can be utilized to illuminate the nature free will will surely be a topic of future exploration.

On the one hand, interventionism seems to be a particularly good approach for those who work on free will. Recall that the question of whether we have free will can be understood as both a metaphysical and ethical question, where the free will or control condition is typically understood as the metaphysical condition necessary for agents to be the appropriate targets of our judgments of moral responsibility, praise, and blame. In other words, free will is a metaphysical relation beholden to our normative practices. Interventionism approaches causation in a similar way. For instance, interventionist approaches require us to choose which variables are endogenous and exogenous in modeling causal relationships (Ismael 2016). If we think that an agent's intentions or the output of their CAS is causally relevant to a given outcome, we can assign independent variables to these features (as opposed to others) in constructing our model. In this way, our interests and values are essential features of interventionist causal models, for they are key to determining how systems are to be modeled. But interventionist approaches to causation are neither purely normative nor subjective. As Ismael explains:

> Which networks we are interested in, and which variables we treat as endogenous and exogenous, are determined by context and purpose… [C]ausal relations… are inductive generalizations from testable regularities and are grounded in the fact that the world is built from a collection of relatively autonomous rule-governed components (Ismael 2016: 136)

Because of the success interventionism has found in other areas of both philosophy and science, and because the interventionist approach to causation is able to bridge the formidable gap between our practices and metaphysics, many working on free will find this approach to be a fruitful and illuminating source of inspiration.

---

[7] Importantly, this soft-line response only works on manipulation arguments that feature an intentional manipulator (Deery & Nahmias 2017, 1273). If Diana thought she'd try her hand at manipulation and got lucky or if a natural event is responsible for Manny being the way he is (as discussed by Pereboom (2014)), then Deery and Nahmias must take the hard-line in response to such cases.

On the other hand, some philosophers are skeptical that interventionism can be used to inform accounts of free will. Sara Bernstein (forthcoming), for example, argues that because interventionism relies on normative practices and human judgment, it provides an account of *causal explanation* but not *causation*, which is a mind-independent feature of the world (Bernstein forthcoming: 15). This can be understood as both an objection to interventionist accounts of causation in general and an objection to their use in developing accounts of free will in particular. The general worry is that interventionist accounts are only able to give us an account of our causal judgments or explanations, but this is importantly different from giving us an account of causation. But there's also a particular worry about the use of interventionist approaches to causation in accounts of free will: if we want the answer to the question 'what is a cause?' to inform the answer to the question 'what is free will?', we don't want our intuitions about the latter question to affect the answer to the former. This would be, if not entirely circular, quite uninformative.

Deery and Nahmias are aware of this objection, at least in its general form. They argue:

> "...[O]bjections claiming that interventionism gives us, at best, an account of our casual practices or judgments, rather than providing (as it should) a theory of what metaphysically causes what, come close to begging an important methodological question…Surely it is an open methodological question…whether the interventionist approach to causation—as well as to other topics in metaphysics—is the correct one to adopt…We do not assume that it is, yet we insist that it is a strong contender, which, if correct, supports a soft-line response to the Manipulation Argument." (Deery & Nahmias 2017: 1270).

Notice that Deery and Nahmias can be right about the methodological point as to whether interventionism counts as an account of causation, but this doesn't address the more particular objection to using interventionist accounts of causation to inform our accounts of free will. While there might be nothing dialectically improper about relying on our intuitions about free will to generate an account of causation, there does seem to be something illicit about relying on our intuitions about free will to generate an account of causation that will then go on to inform an account of free will.

An objection to Deery and Nahmias' soft-line objection may help to illustrate this criticism. Deery and Nahmias consider several objections to their view, including the objection that their causal sourcehood requirement is too demanding. In many cases, there could be causal variables outside of an agent's CAS that could bear a more invariant relation to an agent's action than the output of an agent's CAS. In such cases, the output of an agent's CAS would not be the causal source of the agent's action (or at least not the only causal source). Deery and Nahmias take this to be a positive feature of their view, for it allows them to make sense of free will that comes in degrees and our scalar practices of holding agents morally responsible.[8] In such cases, the agent might still be free and morally responsible for the action in question, though not maximally responsible.

I think that Deery and Nahmias' response to this objection moves too quickly. Diana is the causal source of Manny's stealing the wallet because she *acts through* his compatibilist agential structure, ensuring that he steals the wallet. But our compatibilist agential structures are often acted through, without eliminating or mitigating the degree to which we are free and responsible. We often perform blameworthy actions because others ask us to, or put the idea in our head, or give us an incentive to do so. These kinds of causal variables act through our compatibilist agential structures, and while they

[8] For other interventionist approaches to responsibility that comes in degrees, see Chockler and Halpern (2004) and Halpern (2015).

don't ensure that we'll perform these blameworthy actions, they do make it more likely. In these kinds of cases, Deery and Nahmias' analysis seems to commit them to the claim that these external causal variables are the causal sources of our blameworthy behavior, thus counterintuitively eliminating (or at least reducing) the degree to which the agents who performed the blameworthy actions are free and responsible.

Imagine the following case: A child, who loves ponies very much, asks his mother to get him a pony. The mother, like most parents, doesn't have the financial resources to buy a pony but because she loves her son very much and her son wants a pony very badly, she decides to steal a pony for him. Now imagine that the mother acts on her decision and steals a pony. In this case, it strikes me that the mother acted freely in stealing the pony and is morally responsible for doing so. But is she the causal source of the action?

Using Deery and Nahmias' analysis, we can model the relationship between the son's request for a pony (variable *R*, value r), the mother's decision to steal a pony (variable *D*, value d), and the mother stealing the pony (variable *S*, value s).

$$R=r \rightarrow D=d \rightarrow S=s$$
$$D=d \rightarrow S=s$$

If we hold the relevant background conditions, *C*, fixed, interventions on both *R* and *D* will produce similar changes to the value of *S*. If the son never asked for a pony, the thought of stealing a pony would never have occurred to the mother, and she wouldn't have stolen the pony. And if the mother had never decided to steal the pony, then she wouldn't have committed the crime. But if we consider a range of changes in *C*, *R* bears a stronger invariance relation to *S* than *D* bears to *S*. Because the son asks his mother to give him a pony, we can predict that she would go through with stealing a pony across more changes to the circumstances than if she had simply decided to steal a pony to give to her son. Perhaps the mother would be willing to climb a higher fence, or undergo the crime even in bad weather if her child asked her to get him a pony, but she wouldn't do so if she simply decided to steal a pony for her son without being asked. Recall that when evaluating the relationship between *D* and *S*, we must ignore the contribution of *R*. So, for all the mother knows, her son might not even want a pony! Thus, the son's request for a pony (*R*) bears a stronger invariance relation to the mother's action of stealing the pony (*S*) than the mother's decision alone (*D*) bears to the event. And according to Deery and Nahmias' account, this means that the son's request for a pony is the causal source of the mother's action of stealing the pony, not her decision to do so. But surely the mother is morally responsible for stealing the pony, even if the causal source of the action was her child's request. And while Deery and Nahmias leave open the possibility that this can simply mitigate, as opposed to eliminate, the degree to which the mother is free and morally responsible, it's odd that the mother's degree of moral responsibility is *at all* mitigated simply because she was responding to a child's request. Surely stealing a pony because your child asks you for a pony is just as blameworthy as stealing a pony because you know your child likes ponies.

One natural response to this objection is to argue that in the case described above, the mother's compatibilist agential structure encompasses the fact that her child asked her for a pony. After all, many accounts of free will rely on notions of reasons-responsiveness (Fischer & Ravizza 1998) and reasons sensitivity (Sartorio 2016), so it would make sense to include the reason to engage in the blameworthy behavior as a feature of the agent's compatibilist agential structure. Other accounts rely on first and second-order desires (Frankfurt 1971), and perhaps the child's request for a pony gives rise to a first-order desire to steal a pony that is in-line with a second-order desire to desire things that will make one's child happy. And if this is the case, then even when we're doing an intervention

on *D* and ignoring the role of *R*, we still must take into account that the child asked the mother for a pony (either as a reason for acting or a desire to be fulfilled), in which case *D* would bear at least as invariant a relation to *S* as *R* would. In this case, *D* really would be the causal source of *S*, and Deery and Nahmias could argue that the mother is fully responsible for stealing the pony.

But notice that this response relies on using other views of free will to determine which variables are endogenous to the model. But the very reason we constructed the model was to answer the question: "Was the mother free to, and morally responsible for, stealing the pony?" While there's nothing illegitimate about using views of free will to assign variables in an interventionist model when the question we're trying to answer is about causation, there is something dangerously circular and uninformative about using a conception of free will to determine which variables to represent within an interventionist model when the question we're trying to answer is about free will.

Of course, the above is a very preliminary worry and there's much to say in Deery and Nahmias' defense. Here I only want to argue that as the influence of interventionist models of causation grows, a fruitful topic of research will be how to incorporate interventionism into accounts of free will in an informative and illuminating way.

3. Sartorio's Actual Causal Sequence View

In her recent book *Causation & Free Will* (2016), Carolina Sartorio develops the Actual Causal Sequence view of free will, or ACS. Like Deery and Nahmias, Sartorio takes causation to play a central role in free will. However, unlike Deery and Nahmias, Sartorio doesn't commit to a single account of causation.[9] Rather, Sartorio identifies key features of causation (or a relevantly similar metaphysical relation) and uses them to develop ACS. There isn't enough room to do justice to Sartorio's rich and innovative account of free will. In this chapter, I'd like to focus on one feature in particular—the interplay between the inspiration for the view—Frankfurt cases—and the commitments that come with the features of causation Sartorio uses to develop her view.

Like others who develop actual-sequence views,[10] Sartorio takes the inspiration of ACS to be Frankfurt cases. Frankfurt cases, originally developed by Harry Frankfurt (1969), are designed to illustrate the irrelevance of the ability to do otherwise and the importance of an agent's actual sequence in accounting for free will and moral responsibility. Sartorio develops her own Frankfurt case as follows:

> Frankfurt Case: A neuroscientist has been secretly monitoring the brain processes of an agent, call him Frank, who is deliberating about whether to make a certain choice, C. The neuroscientist can reliably predict the choices that Frank is about to make by looking at the activity in his brain, and can also manipulate Frank's brain in a way that guarantees that Frank will make choice C. He plans to intervene if he predicts that Frank will not make choice C on his own. As it happens, Frank makes choice C on his own, motivated by his own reasons, and without the intervention of the neuroscientist (who correctly predicts that Frank would make that choice on his own). (Sartorio 2016: 13)

Sartorio isolates two intuitions that are generated by the above Frankfurt case:

---

[9] In fact, she leaves open the possibility that freedom is grounded in a quasi-causal or other metaphysical relation that plays a similar role to that of causation (Sartorio 2016: 45).

[10] See compatibilists such as Fischer and Ravizza (1998), McKenna (2008), and incompatibilists such as Pereboom (2001).

> **Intuition 1**: Frank (our agent in a Frankfurt case) is in control of his act despite his lack of robust alternatives.
> **Intuition 2**: What determines whether Frank is in control of his act is how he actually came to perform the act. (Sartorio 2016: 17)

Intuition 1 isolates what *isn't* relevant to free will and control while Intuition 2 isolates what *is*. The fact that, intuitively, Frank is free and morally responsible for making choice C though he couldn't have done other than make choice C counts against views that defend the principle of alternative possibilities. The intuition that the actual sequence of events that lead Frank to choose choice C is what makes him free and morally responsible lends support to actual-sequence approaches to free will. Of course, an incredible amount has been written on whether these intuitions can either successfully undermine (in the case of Intuition 1) or support (in the case of Intuition 2) the relevant philosophical theses. Sartorio doesn't engage in this debate, but her development of the ACS view is inspired by these intuitions (Sartorio 2016: 16).

Sartorio develops two grounding claims that actual-sequence theorists are committed to in virtue of taking Frankfurt cases seriously. She then provides causal interpretations of both:

> **Positive grounding claim**: freedom is grounded in facts about actual causal histories. (Sartorio 2016: 21)
> **Negative grounding claim**: freedom isn't grounded in anything other than actual causal histories. (Sartorio 2016: 28-29)

Sartorio then argues that these two claims together support a further claim about supervenience: "An agent's freedom with respect to X supervenes on those elements of the causal sequence issuing in X that ground that agent's freedom." (Sartorio 2016: 29). According to Sartorio, this claim captures the key insight from Frankfurt cases and actual-sequence theorists of free will should remain loyal to it in developing their views.

Sartorio then goes on to isolate several key features of causation (or a relevantly similar metaphysical relation) that can bolster the grounding and supervenience claims of ACS. In this chapter, I'd like to focus on two of those features:

> OMISSIONS: omissions and other kinds of absences can enter into causal relationships. (Sartorio 2016: 46)
> EXTRINSICNESS: a causal relation between C and E may obtain, in part, owing to factors that are extrinsic to the causal process linking C and E. (Sartorio 2016: 71)

On Sartorio's account, we can be responsible for our omissions, just as we can be responsible for our actions,[11] and whether we cause these omissions and actions can depend on extrinsic factors. EXTRINSICNESS and OMISSIONS interact with each other in interesting ways and help defend ACS against counterexamples to the supervenience claim. Take this pair of cases originally discussed by van Inwagen (1983):

> **Phones:** I witness a man being robbed and beaten. I consider calling the police. I could easily pick up the phone and call them. But I decide against it, out of a combination of fear and laziness.

[11] For another defense of the moral relevance of omissions, see Bernstein (2014, 2016).

> **No Phones:** Everything is the same as in Phones except that, unbeknownst to me, I couldn't have called the police (the phone lines were down at the time). (Sartorio 2016: 56)

Intuitively, we are morally responsible in Phones but not in No Phones. Some, like van Inwagen (1983) argue that this is because we have the ability to do otherwise in Phones but not in No Phones. If van Inwagen is right, this would threaten the supervenience claim Sartorio, and other actual-sequence theorists, defend. But Sartorio is able to explain the asymmetry in terms of the extrinsicness of causation.

In Phones, it's clear that the combination of fear and laziness caused the failure to call the police (because we're assuming that omissions can be caused), according to Sartorio (2016: 69). In No Phones, the combination of fear and laziness, though it caused the failure to *try* to call the police, it didn't cause the failure to call the police. Rather, an extrinsic factor—the fact that the phone lines were down—makes it impossible that the combination of fear and laziness could have caused the failure to call the police (Sartorio 2016: 90). An extrinsic factor is also involved in the causal story to be told in Phones. The combination of fear and laziness caused the failure to call the police in part because the phone lines *were* working (Sartorio 2016: 87). Notice that extrinsic features determine the causal relationships that obtain in both Phones and No Phones, and thus they determine the extent to which the agents in these cases are free (and responsible), but both of these features are out of the agents' control—they are a matter of luck. Sartorio calls this kind of luck that arises from the extrinsicness of causation "Type-2 luck."

> Type-2 Luck: The agent is not in control of some facts external to the causal history that help determine its composition. (Sartorio 2016: 89)

Interestingly, omissions and other kinds of absences are more susceptible to type-2 luck than actions and other positive events; Sartorio calls this Type-2 Luck Asymmetry (Sartorio 2016: 91). This is because extrinsic features are more likely to affect the causal history of omissions than they do actions. To illustrate this asymmetry, Sartorio compares No Phones to a Frankfurt-style version of Phones:

> **Frankfurt-style Phones [Action]:** Again, I witness the man being robbed and beaten. This time the phones are working. With a lot of effort I manage to overcome my fear and laziness, pick up the phone and call the police. Unbeknownst to me, a neuroscientist has been monitoring my brain. Had I wavered in my decision, he would have manipulated my brain in such a way that I would still have made the same choice. (Sartorio 2016: 90)

Both Phones and Frankfurt-style Phones [Action] feature extrinsic factors, but these features don't affect the causal history in these cases symmetrically. The fact that the phone lines are down in Phones is an extrinsic factor that makes a causal difference—because the phone lines are down, the combination of fear and laziness doesn't cause the failure to call the police. In Frankfurt-style Phones [Action], the neuroscientist is an extrinsic factor that doesn't make a causal difference. Overcoming the fear and laziness is a cause of the agent calling the police, regardless of whether a neuroscientist stands by to intervene.

This asymmetry raises many questions, namely: how does ACS handle Frankfurt-style cases that involve omissions? Indeed, Sartorio considers such a case:

> Imagine that I decide on my own not to call the police in circumstances where the phones were working; however, had I hesitated in making that choice, a neuroscientist who had been monitoring my thoughts would have intervened by manipulating my brain and, as a result, I would have made the same choice. Am I free and responsible for not calling the police in this case? (Sartorio 2016: 91)

It strikes me that the agent in the above case was free to, and responsible for, not calling the police, just as Frank is morally responsible in Sartorio's original Frankfurt case. But if one is committed to Type-2 Luck Asymmetry, this is far from clear. Rather than discuss how ACS would address Frankfurt-style cases featuring omissions, Sartorio moves past the issue, citing the vast literature on this very contentious topic and arguing that it's unclear how generalizable the asymmetry between omissions and actions is.[12] She does argue that the extent to which the asymmetry extends will depend entirely on the causal asymmetry (Sartorio 2016: 91).

Sartorio's discussion of Type-2 Luck Asymmetry brings out an interesting tension between the motivation for ACS and the metaphysical assumptions on which it relies. The main, if not sole, motivation for ACS (and presumably other actual-sequence views) is the set of intuitions generated by Frankfurt cases. As Sartorio argues, these intuitions can be captured by two grounding claims, one positive and one negative, which in turn support a claim about supervenience: "No difference in freedom without a difference in the relevant elements of the causal sequence" (Sartorio 2016: 32). Sartorio then defends the supervenience claim from purported counterexamples by relying on the extrinsicness of causation. However, it is this very metaphysical assumption that makes it difficult for ACS to capture our intuitions when it comes to Frankfurt cases that feature omissions. But why should our judgments about a Frankfurt case featuring an omission be beholden to our metaphysical commitments? Why can't they be treated as the motivation for an actual sequence view, like the judgments about Frankfurt cases featuring actions?

We can alter Sartorio's original Frankfurt case to feature an omission:

> **Frankfurt Case-Omission**: A neuroscientist has been secretly monitoring the brain processes of an agent, call him Frank, who is deliberating about whether to make a certain choice, C. The neuroscientist can reliably predict the choices that Frank is about to make by looking at the activity in his brain, and can also manipulate Frank's brain in a way that guarantees that Frank will **not** make choice C. He plans to intervene if he predicts that Frank will make choice C on his own. As it happens, Frank does **not** make choice C on his own, motivated by his own reasons, and without the intervention of the neuroscientist (who correctly predicts that Frank would **not** make that choice on his own).

I take it that this case can generate the intuition that Frank is morally responsible for not choosing choice C just as clearly and forcefully as Sartorio's original case can generate the intuition that Frank is morally responsible for choosing choice C. Notice that our intuitions in response to this pair of cases can still be captured by both Sartorio's positive and negative grounding claims, which can still support the supervenience claim—it can still be true that there's no difference in freedom without a difference in the causal sequences. Of course, we would need to rely on very different features of causation (or a metaphysically similar relation) to defend the supervenience claim in the face of

[12] For discussion of Frankfurt-cases featuring omissions, see: Swenson (2015, 2016), Fischer and Ravizza (1998), Clarke (1994, 2011, 2014).

purported counterexamples. Though we would most assuredly hold onto OMISSIONS**,** we might not be able to rely on EXTRINSICNESS, for example.

Whether we should conceive of Frankfurt cases featuring actions and Frankfurt cases featuring omissions as playing the same dialectical role is an interesting question. While some Frankfurt cases featuring omissions generate intuitions similar to the intuitions generated by Frankfurt cases featuring actions (like Frankfurt Case-Omission), others arguably do not. Indeed, some rely on this intuitive asymmetry to object to actual sequence views of free will (Swenson 2015, 2016). What generates these varying intuitions when it comes to omissions? Is it possible for an actual sequence theorist to come up with an in principle difference between cases like Frankfurt Case-Omission and No Phones? And, more generally, when should we let our intuitions about freedom guide our metaphysical commitments and when should we let metaphysical commitments guide our judgments about freedom? These are difficult questions, and while many philosophers are currently grappling with them now, they are likely to become even more important as the focus on the metaphysics of causation in the free will literature grows more prominent.

4. Bernstein's principle of proportionality

So far I've reviewed two different approaches to the relationship between free will and the metaphysics of causation. Deery and Nahmias utilized a specific view of causation, the interventionist approach, to develop an account of free will that can withstand manipulation objections. Sartorio, rather than utilize a specific view of causation, relied on several common properties of causation to support ACS. Though both approaches are innovative and serve to move the research on free will forward, they each raise interesting and perplexing questions about the relationship between our normative and metaphysical commitments surrounding causation and free will. The use of interventionist causation when attempting to discover the causal source of agents' actions can either produce counterintuitive results or be uninformative, depending on the way in which we conceive of agents' compatibilist agential structures and represent them in our causal models. And while EXTRINSICNESS might be a plausible feature of causation, it leads to counterintuitive results when it comes to Frankfurt cases featuring omissions. While the above authors clearly agree that free will and causation are intimately connected, it's less clear how we can use theories of causation to illuminate the nature of free will. Sara Bernstein drives this worry home in her essay "Causal Proportions and Moral Responsibility" (forthcoming).

Bernstein begins with the intuition that we can only be morally responsible for what we cause (forthcoming: 1). Given this intuition, if one thinks that moral responsibility (or free will) can come in degrees, then it seems the following principle is true:

> **Proportionality**: An agent's moral responsibility for an outcome is proportionate to her actual causal contribution to the outcome. (Bernstein forthcoming, page 3).

On the face of it, this claim is much weaker than the views the authors above defend and can be accepted even by those who defend a variety of views about the relationship between free will and causation, but it too proves difficult for views of free will to fully accommodate. While this principle tells us that agents can be more or less responsible for outcomes, it alone cannot tell us when an agent is more or less responsible than another agent. To answer this question, we would need to know when one agent causally contributes to an outcome to a greater degree than another. But this cannot easily be accounted for by current theories of causation. First, Bernstein argues that depending on whether you favor a productive or dependent theory of causation (Hall 2004), an agent could causally contribute to an outcome to a greater or lesser degree and thus be either more or less responsible than

another agent (Bernstein forthcoming). According to Bernstein, there is a 'semantic indeterminacy' at play in Proportionality and it's not clear which theory of causation we should use to use to resolve it. Next, Bernstein argues that to truly understand Proportionality, we ought to have a precise metaphysical account of what it means for an agent to be more or less of a cause. But again it's not clear that any theory of causation can give us such an account. Bernstein argues: …[T]he relationship between causation and moral responsibility so often used in moral assessment is much trickier than previously imagined. It is also particularly methodologically fraught if current causal theories are to be the guides… (Bernstein forthcoming: 19-20).

I won't delve into the intricacies of Bernstein's argument here. I would like to highlight how difficult the task of incorporating the metaphysics of causation into accounts of free will threatens to be given Bernstein's analysis. In the sections above, I pointed to potential worries involved in relying on interventionist approaches to causation and the extrinsicness of causation. But Bernstein argues that there is *no* (single) theory of causation (or set of metaphysical assumptions about causation) that is able to successfully illuminate the nature of free will.[13]

Bernstein concludes her essay by arguing: "There is much work to be done before we can trust that linking moral responsibility to metaphysical theories of causation clarifies our theories rather than obfuscates our thinking on these matters" (Bernstein forthcoming: 20). But what kind of work is required and who should be given this task, those who work on causation or those who work on free will? Should those developing accounts of free will wait to incorporate theories of causation until there is consensus on what causation amounts to? This doesn't seem particularly promising. Nor does it seem advisable to entirely ignore the role of causation in developing accounts of free will (if you take causation to be relevant to free will). But what work can be done in the absence of a single, determinate theory of causation that can perfectly account for the entirely of our intuitions regarding causation and moral responsibility?

One tentative answer is that those who work on free will can get clear on exactly the kind of work they want a theory of causation to accomplish. Take the two cases Bernstein originally compares in her essay:

> **Victim:** Two independently employed assassins, unaware of each other, are dispatched to eliminate Victim. Being struck by one bullet is sufficient to kill Victim. Each assassin shoots, and Victim dies.
>
> **Hardy Victim:** Two independently employed assassins, each unaware of the other, are dispatched to eliminate Victim. Unbeknownst to both assassins, Victim is particularly hardy, and requires two bullets for his demise. Each assassin shoots, and Victim dies. (Bernstein forthcoming: 1)

The first is a case of overdetermination and the second is a case of joint causation. One can grant that different theories of causation will provide different assessments of causal contribution in these cases, and that there are no rules for which theory of causation we should use to determine who is more causally responsible, as Bernstein argues (forthcoming: 9). But whether this is a problem depends entirely on what we want from a theory of causation.

---

[13] Some have begun to develop accounts of metaphysical relations distinct from causation, like production, that can help ground free will and moral responsibility. See Beckers and Vennekens (forthcoming).

I don't have a clear intuition as to who is more morally responsible for killing Victim, either the assassins in Victim or Hardy Victim. Or rather, I have conflicting intuitions—in one sense the assassins in Victim seem more morally responsible and in another the assassins in Hardy Victim seem more responsible. And the two theories of causation that Bernstein discusses are able to accommodate these conflicting intuitions: on a counterfactual approach, the assassins in Hardy Victim are more causally responsible (and, given Proportionality, morally responsible) and on a productive approach to causation, the assassins in Victim are more causally responsible (and thus morally responsible). If what we want from a theory of causation is to be able to accommodate, explain, or perhaps even ground our judgments of moral responsibility, then we have a success.

But if what we want is for a single theory of causation to be able to resolve all outlier and problem cases of free will and moral responsibility, then we have a failure. But why would we ever expect a single theory of causation to be able to do that? There are outlier and problem cases in the causation literature as well, and if no theory can render coherent all of our causal intuitions, then it's unreasonable to expect such a theory to render coherent all of our intuitions about moral responsibility. Of course, Bernstein's conclusion stands: there's much work to do. But those who work on free will can get their hands dirty right alongside the metaphysicians.

First, Bernstein argues that there are "no clear, principled rules for which type of causal relation should be used" to evaluate cases like Victim and Hardy Victim (forthcoming: page 9). Perhaps the literature on free will and moral responsibility could be helpful in developing such rules (if we want such rules in the first place). While there might be nothing in the causation literature that would lead one to favor one approach to causation over another, there might be normative concerns that could favor one approach. Second, there might be room for both (or more) notions of causation in our accounts of free will and moral responsibility. Perhaps we can be pluralists about causation—different notions of causation can ground different notions of free will and/or responsibility, for example.[14] Of course, there are many ways of being a pluralist about causation, free will, and moral responsibility, and it's far beyond the scope of the chapter to discuss the details of such views. I only want to suggest here that the development of such views might prove to be an important focus of future research.

At the very least, we should adopt a methodological pluralism about causation. After all, it turns out that those working on free will want very different things from an account of causation. Deery and Nahmias want an account of causation that can make sense of the intuition that manipulated agents aren't free and responsible while determined agents are. Sartorio wants an account of causation upon which free will can supervene. And Bernstein seeks an account of causation that could come in degrees and explain our scalar judgments of moral responsibility. It's unlikely that a single conception of causation can accomplish all of these tasks, and perhaps we shouldn't expect it to. Rather, we can make real headway into the nature of free will by adopting many different approaches to causation, at least in the absence of a consensus about the true, single nature of causation.

5. Conclusion

In the concluding section of this chapter, I'd first like to summarize some of the questions that the current work on causation and free will raise:

[14] I've reviewed several distinct conceptions of free will above. Perhaps some of these conceptions of free will can hang together in a single theory. For distinct conceptions of moral responsibility that hang together, see Watson (1996) and Shoemaker (2011).

- When utilizing an interventionist approach to causation in an account of free will, how can we assign variables to our causal models in a way that does not producing counterintuitive results or run the risk of providing circular/uninformative accounts of free will?
- Should those who defend actual sequence accounts of free will be committed to symmetrical judgments in the face of Frankfurt cases featuring omissions and Frankfurt cases featuring actions?
- What work can be done when it comes to free will in the absence of a single, determinate theory of causation that can perfectly account for the entirely of our intuitions regarding causation and moral responsibility?

And here are a few questions that can guide future research:

- In developing accounts of free will that involve causation, must philosophers defend a particular view of causation (like Deery & Nahmias), or simply rely on a set of plausible metaphysical assumptions about causation (like Sartorio)?
- When should we let our intuitions about freedom guide our commitments regarding causation and when should we let these commitments guide our judgments about freedom?
- Can those who work on free will embrace a pluralism about causation in a fruitful way?

Finally, though I focused solely on the relationship between causation and free will in this chapter, notice that these guiding questions for future research apply to all areas in which metaphysics intersects with free will.

- In developing accounts of free will that involve metaphysical concepts, must philosophers defend a particular view of that concept, or simply rely on a set of plausible metaphysical assumptions about it?
- When should we let our intuitions about freedom guide our metaphysical commitments and when should we let these commitments guide our judgments about freedom?
- Can those who work on free will embrace a pluralism about any/all relevant metaphysical concepts in a fruitful way?

Works Cited

Beckers, S. and Vennekens, J. (forthcoming) "A Principled Approach to Defining Actual Causation," *Sythese*.

Bernstein, Sara (2014) "Omissions as Possibilities." *Philosophical Studies*. 167. 1, pp. 1-23.

Bernstein, Sara (2016) "Omission Impossible." *Philosophical Studies*. 173. 10, pp. 2575-2589.

Bernstein, Sara (forthcoming) "Causal Proportions and Moral Responsibility." In Shoemaker, D. (ed.) *Oxford Studies in Agency and Responsibility*. Oxford: Oxford University Press.

Campbell, John (2010) "Control Variables and Mental Causation." *Proceedings of the Aristotelian Society*. 110. pp. 15-30.

Chockler, Hana and Halpern, Joseph (2004) "Responsibility and Blame: A Structural-Model Approach." *Journal of Artificial Intelligence*. 22. pp. 93-115.

Clarke, Randolph (1994) "Ability and Responsibility for Omissions." *Philosophical Studies*. 73. 2, pp. 195-208.

Clarke, Randolph (2011) "Omissions, Responsibility, and Symmetry." *Philosophy and Phenomenological Research*. 82. pp. 594-624.

Clarke, Randolph (2014) *Omissions: Agency, Metaphysics, and Responsibility*. Oxford: Oxford University Press.

Deery, Oisín and Nahmias, Eddy (2017) "Defeating Manipulation Arguments: Interventionist Causation and Compatibilist Sourcehood." 174. 5, pp. 1255-1276.

Fischer, John Martin and Ravizza, Mark (1998) *Responsibility and Control*. Cambridge: Cambridge University Press.

Frankfurt, Harry (1969) "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy*. 66. 23, pp. 829-839.

Frankfurt, Harry (1971) "Freedom of the Will and the Concept of a Person." *Journal of Philosophy*. 68. 1, pp. 5-20.

Hall, Ned (2004) "Two Concepts of Causation." In Collins, J., Hall, N., and Paul, L. A. (eds.) *Causation and Counterfactuals*. Cambridge, MA: MIT Press.

Halpern, Joseph (2015) "Cause, Responsibility and Blame: A Structural-Model Approach." *Law, Probability & Risk*. 14. 2, 91-118.

Ismael, J. T. (2013) "Causation, Free Will, and Naturalism." In Kincaid, H., Ladyman, J., and Ross, D. (eds.) *Scientific Metaphysics*. pp. 208-235. New York: Oxford University Press.

Ismael, J. T. (2016) *Why Physics Makes Us Free*. New York: Oxford University Press.

Kane, Robert (1996) *The Significance of Free Will*. New York: Oxford University Press.

McKenna, Michael (2008) "A Hard-line Reply to Pereboom's Four-case Argument." *Philosophy and Phenomenological Research*. 77. 1, pp. 142-59.

Mele, Alfred (2013) "Manipulation, Moral Responsibility, and Bullet Biting." *Journal of Ethics*. 17. 3, pp. 167-184.

O'Connor, Timothy (1995) *Agents, Causes, and Events: Essays on Indeterminism and Free Will*. New York: Oxford University Press.

Nelkin, Dana (2011) *Making Sense of Freedom and Responsibility*. New York: Oxford University Press.

Pearl, Judea (2009) *Causality*. Cambridge: Cambridge University Press.

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

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

Roskies, Adina (2012) "Don't Panic: Self-authorship Without Obscure Metaphysics." *Philosophical Perspectives*. 26. 1, pp. 323-342.

Sartorio, Carolina (2016) *Causation & Free Will*. Oxford: Oxford University Press.

Shoemaker, David (2011) "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility," *Ethics*. 121. 3, pp. 602-632.

Swenson, Philip (2015) "A Challenge for Frankfurt-style Compatibilists." *Philosophical Studies*. 172. 5, pp. 1279-1285.

Swenson, Philip (2016) "The Frankfurt Cases and Responsibility for Omissions." *The Philosophical Quarterly*. 66. 264, pp. 579-595.

Van Inwagen, Peter (1983) *An Essay on Free Will*. New York: Oxford University Press.

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

Watson, Gary (1996) "Two Faces of Responsibility," *Philosophical Topics*. 24. 2, pp. 227-248.

Wolf, Susan (1990) *Freedom Within Reason*. New York, NY: Oxford University Press.

Woodward, James (2003) *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press.