Review of Carolina Sartorio's *Causation and Free Will*

Sara Bernstein

Carolina Sartorio's *Causation and Free Will* is the most important contribution to the free will debate in recent memory. It is innovative and rigorous, and makes genuine progress on the classic long-standing philosophical problem of whether or not we are free, and if so, in what this freedom consists.

Whether or not we are free depends only on facts about the actual world rather the availability of alternative possibilities, argue proponents of *actual sequence views* of free will. Actual sequence views are dialectically situated against views that hold that freedom is a matter of ability to do otherwise. This latter sort of view has been less popular since the advent of Frankfurt cases, which purport to show that agents can act freely and be held responsible for their actions without having the ability to do otherwise.

Nonetheless, the details of actual sequence views have proven difficult to articulate and defend. Many actualists sneak modal concepts (possibilities, counterfactuals, dispositions) into their views. Other forms of actualism are overly schematic and do not give the particular details of the "actual sequence" in question.

Carolina Sartorio fruitfully enters the debate at this juncture. In *Causation and Free Will*, Sartorio uses the metaphysics of causation in the service of fleshing out an actual sequence view of free will, with extremely significant results. We are free, Sartorio holds, only if "our acts have the right kinds of causes." (p. 3) More specifically, "all that matters to the freedom of an act is how the agent came to perform the act, or the **actual history** of the act" (p. 18) (emphasis added). The actual history of the act is explicated in terms of the metaphysics of causation.

After laying the groundwork for the actual sequence view in Chapter 1, Chapters 2 and 3 turn to the metaphysics of causation, the details of which constitute the core of Sartorio's actual sequence view. First, Sartorio is committed *absence causation*, the idea that absences (such as a failure to water one's plant) can be causes, effects, and causal intermediaries. Absences are subjects of free will: we are free to fail to do things as well as free to do things. Second, Sartorio is committed to the idea that causation is in some cases *intransitive*: if *c*1 causes *c2*, which thereby causes *c*3, *c*1 is not necessarily a cause of *c*3. Third, Sartorio is committed to the extrinsicality of the causal relation: whether or

not *c* is a cause of *e* sometimes depends on factors outside of the causal process between *c* and *e*. Finally, Sartorio denies the centrality of counterfactual dependence to causation, instead holding that causation is partially a *difference-making* relation, similar to but not exactly like counterfactual dependence. Of difference-makers, the idea is that "their effects wouldn't have been caused by the absence of their causes." (p. 94)

A core piece of Sartorio's view is that there is no difference in freedom without a difference in the relevant elements of the actual causal sequence. This supervenience claim poses a problem, however: there are cases in which agents' sequences of actions and omissions do not seem to differ, and yet there is a difference in freedom. The intuitive idea is brought out in a case Sartorio (drawing on van Inwagen) calls No Phones, in which a bystander considers calling the police when hearing a man being beaten. The bystander decides against it, out of a combination of fear and laziness. Unbeknownst to the bystander, she couldn't have called the police (the phone lines were down at the time). Contrast this case with an identical one in which the phone does work. The bystanders' causal histories are the same, but their freedom with respect to placing the call seems different.

Sartorio's defense is to deny the sameness of causal histories. In the case of the non-working phone, the failure of the phone call to be made is not caused by the decision not to make the call. In the case of the working phone, the failure *is* caused by the decision. Supervenience of freedom upon actual causal sequences is saved, because the causal histories are different. While traditionally such cases are thought to show that it is the nonexistence of alternative possibilities that grounds the lack of free will, Sartorio's central idea is that the failure to make the phone call in the phones-not-working case fails to be free because it does not make the right kind of difference (even if the bystander believes she is making a difference.)

One outcropping of this difference in causal sequences is that Frankfurt cases and their contrast cases lacking interveners are vulnerable to differences in causal sequences as well. In a case lacking a counterfactual intervenor, for example, an agent's deliberation causes the outcome to occur. But in a case with an intervenor waiting in the wings, the agent's deliberation does not cause the outcome to occur, since the agent's deliberation does not make the right kind of difference. If this is true, then agents in

Frankfurt cases are not free, *contra* Sartorio's desired view. Sartorio's solution to this problem is to hold that the agent's deliberations *do* cause the outcome in a Frankfurt case, as the deliberations make a difference that their absence would not have made.

Chapter 4 explicates reasons-sensitivity in terms of the actual causes of an action. Roughly, *reasons-sensitivity* is the trait of responding to reasons in the right way: when I eat with a fork rather than with my hands because it is conventional to do so, I am doing so because I am sensitive to the reasons to do such a thing. Formulating reasons sensitivity non-modally is tricky, since usually responsiveness to reasons is hashed out in terms of dispositions, counterfactuals, possible worlds, and the like. Sartorio's formulation bypasses modal concepts by including absences of reasons in sets of reasons, resulting in the following view:

> *(Causal Reasons-sensitivity)*: An agent is reasons-sensitive in acting in a certain way when the agent acts on the basis of, perhaps in addition to the *presence* of reasons to act in the relevant way, the *absence* of sufficient reasons to refrain from acting in that way, for an appropriately wide range of such reasons. (p. 132)

The inclusion of absences makes a difference insofar as actions can counterfactually depend on absences of reasons in addition to reasons. For example, my attending a faculty meeting depends not just on positive reasons to do so (such as my desire to fulfill professional obligations); it also depends on the absence of reasons not to do so (such as the absence of a five million dollar offer to miss the meeting).

Sartorio's contribution to this debate is original and extremely substantive, and her plethora of ready responses to objections is admirable. Here I will focus on a few of the view's downsides, which do not detract from the importance of Sartorio's view or approach.

First, Sartorio's view is commited to profligate causation by absences of reasons. It is well-known that a commitment to absence causation is vulnerable to the *problem of profligate omissions:* not only does a plant's death counterfactually depend on my failure to water it, but it also counterfactually depends on the Queen of England's failure to water it. Similarly, a commitment to absences of reasons in actual causal sequences is a *de facto* commitment to many more absences of reasons than are intuitively relevant. I

presently do not have a private jet waiting to whisk me away to Barcelona for a sunny vacation, and this absence constitutes one reason among many that I am continuing to write a book review. However, such a reason and ones like it do not seem salient to the actual causal sequence leading to me to write the review.

Second, Sartorio's actual sequence view sneaks in modal concepts via the difference-making characterization of causation.[1] Absences of reasons are causes because "their effects wouldn't have been caused by the absence of their causes." (p. 94) This is a straightforwardly counterfactual claim: an effect would not have been caused in the absence of the absence of a reason. Since difference-making is understood modally, we cannot cordon off the causes from the non-causes without resorting to counterfactuals. Alas, the view is not as free from modal features as it claims to be.

Third, Sartorio's view requires the acceptance of several independently controversial, and at times unpalatable, theses about causation. Foremost among these is the denial of the transitivity of causation. Although there might be independent reasons for doing so (for example, we can deny that Joe's placing the bomb under Jane's desk, which she then defuses, causing her survival), we lose out on the explanatory power and intuitive thrust of the transitivity of causation more generally. This is a particularly steep cost given that the transitivity of causation plays a significant role in moral responsibility for outcomes: often, an agent is morally responsible for an outcome that she caused by causing its cause. And even if the claim is that only some causal sequences are intransitive, such a commitment incurs a new explanatory burden in drawing a non-ad hoc distinction between transitive and intransitive cases of causation.

Finally, there is an independent but related debate about what, exactly, absences are. Sartorio treats absences of reasons as sorts of absences simpliciter. But there is an obvious question about whether this is so: is my failure to water the plant really the same thing as Frank's absence of a reason to shoot Furt? An absence is generally taken to be a broad, type-level event, whereas an absence of a reason is, one assumes, a specific token level mental event. Arguably, absences are at least partially located at or connected to actual-world events. For example, my failure to water the plant is connected to some time and/ or place in the actual world when I could have been doing it. But to what is the

[1] Tognazinni (2016) expresses a similar concern in his review of the book.

absence of a reason connected? A particular brain event? I do not think that absences are metaphysically problematic. And I do not think that absences of reasons are metaphysically problematic. But I do suggest that the two are different beasts, and further elucidation of the latter is necessary for a full-blooded account of free will in terms of actual sequence causation.

The free will debate has, until now, proceeded largely independently of philosophical results on recent developments in the metaphysics of causation. *Causation and Free Will* is a strong contribution to the literature due not just to its content, but also its methodology. Bringing the details of the metaphysics of causation to bear on the problem of free will is a novel approach that will hopefully bear out similarly fruitful results with the careful consideration and application of the metaphysics of properties, relations, modality, and mental causation. Those interested in free will must read this book, which points the way to the future methodology and content of this and similar debates.