

Diana and Ernie return: on Carolina Sartorio's *Causation and Free Will*

Alfred R. Mele¹

Published online: 17 February 2018

© Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract In the final chapter of her *Causation and Free Will*, Carolina Sartorio offers (among other things) a novel reply to an original-design argument for the thesis that determinism is incompatible with free will and moral responsibility, an argument that resembles Alfred Mele's zygote argument in *Free Will and Luck*. This article assesses the merits of her reply. It is concluded that Sartorio has more work to do if she is to lay this style of argument to rest.

Keywords Compatibilism · Free will · Moral responsibility · Original-design arguments · Zygote argument

Carolina Sartorio's *Causation and Free Will* (2016) develops a compatibilist actual-sequence view of free will. The view is very attractive, at least from a compatibilist perspective, and I expect the book to be highly influential. In constructing her view, arguing for it, and defending it against objections, Sartorio attends instructively to such much-discussed topics as Frankfurt-style cases, reasons responsiveness, and various well-known arguments for incompatibilism. But she breaks free from the pack by paying far more attention to the metaphysics of causation than anyone else has done in the literature on free will, and the result is a novel compatibilist position.

When thinking about how to proceed in this article, I contemplated various options, including composing a general discussion of some of the book's major themes and focusing on a single issue instead. I chose the latter path. In her final chapter, Sartorio replies to three different kinds of challenge to her view: ultimacy

✉ Alfred R. Mele
almele@fsu.edu

¹ Department of Philosophy, Florida State University, 151 Dodd Hall, 641 University Way, Tallahassee, FL 32306-1500, USA

arguments, direct arguments, and manipulation arguments. My topic is her reply to the manipulation argument that receives the most attention in that chapter.

1 A goddess and a lightning strike

In Mele (2006), I presented a challenge to compatibilists about free will and moral responsibility that revolved around a story about a goddess, Diana, and a human being she created, Ernie (pp. 187–195). (Readers should not infer I am an incompatibilist about these issues. I was then—and still am—agnostic about compatibilism.)¹ Here is Sartorio’s version of that story.

Diana: Mary is about to use in vitro fertilization to conceive a child. A goddess, Diana, intervenes by inducing a certain genetic mutation in the material in the dish, which she knows will result in the conception of a child, Ernie, with certain innate genetic dispositions. Diana knows that, partly due to those innate dispositions, thirty years later Ernie will perform a terrible act: he’ll murder his uncle to inherit a fortune. Diana induces the relevant genetic mutation because she wants Ernie to murder his uncle, has full knowledge of the state of the world and the deterministic laws, and can predict exactly how the world will evolve after she intervenes in that way. Still, when Ernie murders his uncle thirty years later, he satisfies all of the standard compatibilist conditions on freedom with respect to that act. (p. 159)

Sartorio considers the following argument.

Diana Argument (Simple Version):

- 1) Ernie’s murdering act is not free.
- 2) Ernie meets all of the standard compatibilist conditions when he commits the murdering act.
- 3) Therefore, all standard forms of compatibilism fail. (p. 160)

She reports that although she finds (1) “intuitively plausible,” she “will suggest that there are reasons to be skeptical of the reliability of our intuitions in scenarios of this kind” (p. 160).

Central to Sartorio’s development of this suggestion is the following story.

Lightning Strike: This time there is no goddess Diana. Instead, a flash of lightning strikes the laboratory while the in vitro procedure is taking place, and this results in the same genetic mutation in the material in the dish as the one induced by Diana in the Diana scenario. Everything else is the same as in the Diana scenario. (pp. 160–61)

She reports that her “intuitions change radically” when she considers scenarios of this kind (p. 161); she has “compatibilist-friendly intuitions” about “Lightning

¹ By “compatibilism” in this article, I mean compatibilism both about free will and about moral responsibility, except when I have occasion to be more specific.

Strike" (p. 162). Sartorio also reports that "upon reflection, it seems quite obvious that the source of the genetic mutation could not plausibly make a difference to Ernie's freedom and responsibility: either he is free and responsible in both scenarios, or he is in neither" (p. 161). So which is it?

Describing a strategy used by Michael McKenna (2008, 2014) and John Fischer (2011), Sartorio writes: "instead of using incompatibilist intuitions about manipulation cases to draw incompatibilist conclusions about ordinary cases, one could use compatibilist intuitions about ordinary cases to draw compatibilist conclusions about manipulation cases" (p. 164). Their aim here, as she says, is "to cast doubt on" the move from incompatibilist intuitions to incompatibilist conclusions (p. 164). And one may try to use "Lightning Strike" for the same purpose.

A bold compatibilist may argue as follows.

Bold Compatibilist Argument:

B1. Ernie's murdering act is free in "Lightning Strike" and he is morally responsible for it.

B2. "The source of the genetic mutation could not plausibly make a difference to Ernie's freedom and responsibility: either he is free and responsible in both scenarios, or he is in neither" (Sartorio, p. 161).

B3. Therefore, Ernie's murdering act is free in "Diana" and he is morally responsible for it.

A less bold compatibilist may argue that this bold argument is no less compelling than the following incompatibilist argument and contend that this point results in a stalemate.

Reverse Argument:

R1. Ernie's murdering act is not free in "Diana" and he is not morally responsible for it.

B2. (As above.)

R3. Therefore, Ernie's murdering act is not free in "Lightning Strike" and he is not morally responsible for it.

R1 and B1 are supposed to be based on intuitions. Who has the intuition expressed in R1? Incompatibilists who are in a position to have intuitions about the story presumably find R1 intuitive. (Some incompatibilists may be capable of having only a wholly theory-driven reaction to the story. Such a reaction typically is not counted as an intuition.) Some compatibilists also find R1 intuitive, Sartorio being a case in point. And if some compatibilists find R1 intuitive, we should not be surprised if some agnostics about compatibilism do so as well.

Who has the intuition expressed in B1? Sartorio, a compatibilist, says that she does. But it is unlikely that people who are not compatibilists would have that intuition. Imagine an agnostic about compatibilism, Agnes, who has not encountered original-design stories like "Diana." We present her with "Lightning Strike," and she proceeds to reflect on it. Her responding as follows would be baffling: "What a powerful story! Although I had been agnostic about compatibilism, zooming in on the lightning strike and the effect it has on Ernie's life really helps

me see that he is morally responsible for the murder and kills his victim freely, despite the fact that his world is deterministic.” It is very difficult to see how, in ordinary circumstances, a reflective person who is agnostic about compatibilism would be moved by this story to have a compatibilist-friendly intuition about it. And if that is true of agnostics, it is true as well of incompatibilists.

Agneta is another agnostic about compatibilism who has not encountered original-design stories. We present her with “Diana,” and she reflects on that story. Her reaction resembles Agnes’s: “What a powerful story! Although I had been agnostic about compatibilism, zooming in on what Diana did and the effect it has on Ernie’s life really helps me see that he is not morally responsible for the murder and does not kill his victim freely.” This reaction from an agnostic is far from baffling.

What is my point? The two stories are not on equal footing on the intuition-generating front. First, the details about the lightning strike in no way motivate the intuition that Ernie freely kills and is morally responsible for that action. If anything in “Lightning Strike” motivates that intuition (intuition *FMR*, for short), it is Ernie’s satisfying “all of the standard compatibilist conditions on freedom [and moral responsibility] with respect to that act” (Sartorio, p. 159). And incompatibilists and agnostics who know the relevant literature either reject the claim that satisfying these conditions is sufficient for acting freely and morally responsibly or are agnostic about this claim. Second, the details about Diana’s conduct do motivate, in people who have it, the intuition that Ernie unfreely kills and is not morally responsible for the killing (intuition *NOT*, for short). Third, compare the audience in which “Diana” can reasonably be expected to generate intuition *NOT* with the audience in which “Lightning Strike” can reasonably be expected to generate intuition *FMR*. The former audience includes incompatibilists who are capable of having intuitions on this matter, agnostics, and even some compatibilists. The latter includes neither incompatibilists nor agnostics (with the possible exception of members of either group who are moved to give a compatibilist-friendly response by a set of alleged compatibilist sufficient conditions for free and morally responsible action presented in the story).

It does not follow from these points that “Lightning Strike” cannot be used in an effective rebuttal of the Diana Argument. That issue is still open for discussion. But we have seen some cause for worry about this. If, as I put it, “Diana” and “Lightning Strike” are not on equal footing on the intuition-generating front, stalemating the Diana Argument in the way I mentioned (i.e., by contending that the Bold Compatibilist Argument is no less compelling than the Reverse Argument) will be more difficult than it might have been.

Sometimes authors craft arguments with the intention or hope of persuading every reasonable person who reads them. I certainly have never had that hope for what I called the “zygote argument” (Mele 2006, p. 189), an older relative of the Diana Argument that is based on my own story about Diana and Ernie. As I made clear in Mele (2006), I do not expect the story to move seasoned compatibilists to reject compatibilism (pp. 190–193), and I myself am not persuaded by the zygote argument (pp. 193–194). I am not convinced that Ernie is not morally responsible for his actions and does not act freely. (But I also am not convinced that he does act freely and morally responsibly.) So why did I spin my story about Diana and Ernie

and offer the zygote argument for consideration? Because the story and argument seem to me to help crystallize an incompatibilist worry about compatibilism (see McKenna 2009, pp. 25–26), because I hoped to prompt instructive compatibilist responses to that worry, and because the story and argument might shed some light on why someone might be agnostic about compatibilism, as I am.

How instructive is Sartorio's use of "Lightning Strike" in her response to the Diana Argument? It seems that the story is particularly useful for a segment of her compatibilist audience—namely, those compatibilists who, like her, find it "intuitively plausible" that "Ernie's murdering act is not free" (p. 160) in "Diana" and do not trust that intuition. These troubled or conflicted compatibilists may be pleased by the offer of a good reason to reject the intuition at issue. "Lightning Strike" is very useful in this connection. A predictable compatibilist reaction to the story when combined with the belief that "the source of the genetic mutation could not plausibly make a difference to Ernie's freedom and responsibility" (Sartorio, p. 161; see B2 above) gives these compatibilists a basis for rejecting their intuition that Ernie is unfree in "Diana." But, of course, this basis for rejecting that intuition is not available to people who lack the belief that Ernie's murdering act is free in "Lightning Strike," including ordinary incompatibilists and agnostics. The use I identified for the story in connection with the compatibilist population at issue depends on their believing that Ernie acts freely in "Lightning Strike"; people who do not accept this claim about Ernie—that is, those who disbelieve it and those who are agnostic about it—cannot benefit from the story in the same way. In a way, then, Sartorio seems to be preaching to the compatibilist choir—and, in fact, only to part of it, those who have the intuition that Ernie unfreely murders his uncle in "Diana" and regard it as suspect.

2 A debunking effort

Sartorio's discussion of the Diana Argument has an interesting speculative aspect that I have not yet mentioned. She asks (p. 167): "Why do we tend to feel that Ernie's freedom and responsibility are undermined when the source of the genetic manipulation is intelligent design, but not when it's blind forces?" And she offers an answer, reporting that she will go with her "best guess" (p. 167).

Before I discuss her answer, I comment briefly on the question. What does "we" refer to here? Some people have the feeling Sartorio mentions, and some do not. And some people will have the feeling that Ernie is not free and responsible in "Lightning Strike" whereas others will not. So I read "we" as shorthand for the people to whom Sartorio's question applies, including Sartorio herself, of course.

Sartorio suggests that "something like a psychological 'dilution of responsibility' effect may be at play in these cases" (p. 167). After explaining the effect—the *DoR effect*, for short—she asserts that, if it exists, it "might explain why we tend to think that Ernie is not responsible in the Diana case but he is in the Lightning Strike case" (p. 168). The passage continues as follows:

For, whereas there is no one else to blame for Ernie's act in the Lightning Strike case, there is in the Diana case: Diana herself. Moreover, given that the story is told from the perspective of Diana, with special emphasis on what she does and why she does it, Diana's contribution is particularly salient. In this respect, Diana is like one of those multi-millionaires [mentioned earlier in illustrating the DoR effect] that we're tempted to blame first for not helping, when we're tempted to blame someone for that. Our being particularly focused on Diana might help explain why we are less inclined to see Ernie as a responsible agent. (pp. 168–169)

There might be something to this, and experimental philosophy studies can be conducted to test Sartorio's hypothesis. But what percentage of philosophers who have published on my zygote argument and claimed or argued that Ernie in my story (which is very similar to "Diana") does not freely do what he does and is not morally responsible for doing it would find it plausible that their judgment is explained by the DoR effect—where that effect is understood as the manifestation of a *bias*, of course?² One can survey these philosophers with the aim of answering this question, but my prediction is that the answer would be very close to zero and that these philosophers would not throw in the towel at this point. If one wants to persuade them to abandon their belief about Ernie, appealing to the DoR effect is unlikely to help much. However, compatibilists who share Sartorio's intuitions about Ernie and regard their intuition about Ernie in the Diana stories as suspect may be happy to chalk that intuition up to the DoR effect. So perhaps it is appropriate to mention something else from psychology here—the confirmation bias. Philosophers who are inclined to believe that their intuition that Ernie is unfree in the Diana stories is misleading may be attracted to the hypothesis that the DoR effect explains their intuition, whereas philosophers who believe that their own unfreedom intuition about Ernie is veridical may be attracted to the hypothesis that their intuition is not explained by this effect. If the confirmation bias is at work, it might help to explain why these groups tend to end up where they do. (Compare this speculative, tongue-in-cheek thought with Sartorio's speculative thought about the DoR effect, as it applies to philosophical experts. Is either more worthy of endorsement than the other?)

Be all this as it may, what would seem to be especially useful is for people who believe that Ernie is free and morally responsible in the Diana stories (Sartorio's and mine) to explain what they take to justify that belief and for people who hold the opposite view to explain what they take to justify their own belief on the matter. The rest of us can do our best to assess the opposing arguments (as can people with the opposing views, of course; and if they are vigilant, they might avoid the confirmation bias).

² See, for example, Barnes (2015), Deery and Nahmias (2017), Schlosser (2015), Todd (2013, 2017) and Waller (2014). I mention my story here (Mele 2006) rather than Sartorio's because mine has been in existence long enough to have prompted many published reactions and Sartorio's appeared too recently to have done so.

In this connection, those philosophers who have published on the zygote argument and have claimed or argued that Ernie (in my Diana story) does not freely do what he does and is not morally responsible for doing it can try to counter the suggestion that their judgment is explained by the DoR effect by presenting compelling grounds for the view that they are right about Ernie. And compatibilists who maintain that Ernie does act freely and morally responsibly (in that story)—some of whom may find it *intuitive* that he does—might try to explain why their claims about sufficient conditions for free and morally responsible action are more worthy of our acceptance than is the judgment that Ernie is unfree and lacks moral responsibility.

There is another group of compatibilists, by the way. They believe that Ernie does not act freely and morally responsibly in my Diana story but contend that this does not falsify promising compatibilist sufficient conditions for free action and moral responsibility (Barnes 2015; Deery and Nahmias 2017; Schlosser 2015; Waller 2014). Here is the premise of the zygote argument that they reject: “Concerning free action and moral responsibility of the beings into whom the zygotes develop, there is no significant difference between the way Ernie’s zygote comes to exist and the way any normal human zygote comes to exist in a deterministic universe” (Mele 2006, p. 189). By a “significant” difference I mean one that warrants asymmetrical judgments about Ernie and (some of) the others on the topics of free and morally responsible action. Just as Sartorio contends (see B2 above) that “the source of the genetic mutation” in her stories about Diana and the lightning strike “could not plausibly make a difference to Ernie’s freedom and responsibility: either he is free and responsible in both scenarios, or he is in neither” (p. 161), this premise of the zygote argument asserts that the difference between coming into existence in the way Ernie does and coming into existence in a way human beings normally do cannot plausibly make a difference to the freedom and responsibility of the agents at issue. And the compatibilists at issue now are likely to reject B2. So it seems that Sartorio should be prepared to offer an argument for B2. Her claim that “upon reflection, [B2] seems quite obvious” (Sartorio, p. 161) will be challenged by some compatibilists.

In Mele (2006), I suggested that a defense of the zygote argument’s no-significant-difference premise “might begin with the question how it can matter for the purposes of freedom and moral responsibility whether, in a deterministic universe, a zygote with *Z*’s exact constitution was produced by a supremely intelligent agent with Diana’s effective intentions or instead by blind forces” (p. 190). In response to this question, I wrote:

Imagine a deterministic universe U^* that is a lot like the one at issue, U , but in which Z comes into being in Mary in the normal way and at the same time. It is conceivable that, in U^* , throughout his life, Ernie does exactly what he does in U , down to the smallest detail. Suppose that this is so in U^* . Then, a proponent of the zygote argument might contend that, given the additional facts that, in both universes, Ernie has no say about what causes Z , no say about the rest of the universe at that time, and no say about what the laws of

nature are, the cross-universe difference in what caused Z does not support any cross-universe difference in freedom or moral responsibility. (p. 190)

Sartorio may be thinking along similar lines about the difference between “Diana” and “Lightning Strike.” But it is safe to say that this line of reasoning did not persuade some compatibilists. To bring these compatibilists under her tent—when it comes to the Ernies, that is—with the strategy she employed she will need to find a way to persuade them that B2 is true. (I should add that B2 strikes me as very plausible.)

Elsewhere, I have suggested that *incompatibilist* believers in free will and moral responsibility should try to do more than they have to explain how it is possible for human agents in indeterministic worlds who satisfy their requirements for free action and moral responsibility to act freely and morally responsibly (Mele 2013, 2017). In making this point, I highlighted an analogy with defenses and theodicies in the philosophy of religion. Defenses are critiques of arguments designed to show that the pain and suffering in the world are incompatible with the existence of a perfect God; theodicies are attempts to explain why a perfect God would allow all the bad stuff. Incompatibilist believers in free will and moral responsibility tend to concentrate on rebutting arguments against their position. This is understandable. But the more difficult project of explaining how free action and moral responsibility are possible when typical specifically incompatibilist requirements for them are satisfied is an important one. A similar observation applies to *compatibilist* believers in free will and moral responsibility. They too tend to concentrate, understandably, on rebutting arguments against their position. But convincing arguments for such things as compatibilism itself (about free will and moral responsibility) or the truth of the claim that Ernie (in Sartorio’s Diana story or mine) acts freely and morally responsibly would certainly be desirable.

I myself would love to see an argument that convinces me that Ernie acts freely and morally responsibly in these Diana stories and moves me to abandon my agnosticism in favor of compatibilism. Sartorio has not given me such an argument. But I cannot fault her for that. No one else has given me such an argument either, and showing that compatibilism is true or that Ernie acts freely and morally responsibly in “Diana” is not among the goals of her excellent book. Now, if Sartorio had offered an argument for the claim that Ernie acts freely and morally responsibly in “Lightning Strike,” perhaps things would have gone differently. Why so? If she had offered such an argument, it might have convinced me. And, like her, I believe that B2 is true. So I would have all I need to be confident in the soundness of what I called the “Bold Compatibilist Argument,” the conclusion of which is that Ernie’s murdering act is free in “Diana” and he is morally responsible for it.

The weak link in the Bold Compatibilist Argument, in my opinion, is its first premise, (B1)—the claim that Ernie’s murdering act is free in “Lightning Strike” and he is morally responsible for it. Only a compatibilist would assent to B1 right off the bat; anyone else would need to be talked into it. (Recall that the story takes place in a deterministic world.) Now, Sartorio does not offer the Bold Compatibilist Argument, but B1 does play an important role in her response to the Diana Argument. I have suggested both that incompatibilists and agnostics will find that

response considerably less attractive than compatibilists do and that, within the compatibilist choir, there are those who would reject another central plank in her response—namely (B2), her no-significant-difference claim about the source of the genetic mutation in “Diana” and “Lightning Strike.” Given that B1 and B2 both feature in Sartorio’s response to the Diana Argument (even though that response does *not* include the Bold Compatibilist Argument), my suggestions about them indicate that Sartorio has more work to do if she is to lay the Diana Argument to rest. Soothing the worries of compatibilists who have the intuition that Ernie does not act freely and morally responsibly in “Diana” and would like to see grounds for deeming that intuition misleading is one thing; persuading everyone else that the Diana Argument is a failure is another.

3 Closing comment

Causation and Free Will is a very impressive book. It is engaging, it develops a novel view, and it makes a considerable advance on the compatibilist front. I believe that if I were a compatibilist, I would adopt Sartorio’s view as my own.

Acknowledgements I am grateful to Gabriel DeMarco and Carolina Sartorio for comments on a draft of this article. This article was made possible through the support of a grant from the John Templeton Foundation. The opinions expressed here are my own and do not necessarily reflect the views of the John Templeton Foundation.

References

- Barnes, E. (2015). Freedom, creativity, and manipulation. *Noûs*, 49, 560–588.
- Deery, O., & Nahmias, E. (2017). Defeating manipulation arguments: Interventionist causation and compatibilist sourcehood. *Philosophical Studies*, 174, 1255–1276.
- Fischer, J. (2011). The zygote argument remixed. *Analysis*, 71, 267–272.
- McKenna, M. (2008). A hard-line rely to pereboom’s four-case manipulation argument. *Philosophy and Phenomenological Research*, 77, 142–159.
- McKenna, M. (2009). Compatibilism; State of the Art. In E.N Zalta (Ed.) *The Stanford encyclopedia of philosophy* (winter 2009 edition). <https://plato.stanford.edu/entries/compatibilism/supplement.html>. Accessed 16 Feb 2018.
- McKenna, M. (2014). Resisting the manipulation argument: A hard-liner takes it on the chin. *Philosophy and Phenomenological Research*, 89, 467–484.
- Mele, A. (2006). *Free will and luck*. New York: Oxford University Press.
- Mele, A. (2013). Moral responsibility and the continuation problem. *Philosophical Studies*, 162, 237–255.
- Mele, A. (2017). *Aspects of agency: Decisions, abilities, explanations, and free will*. New York: Oxford University Press.
- Sartorio, C. (2016). *Causation and free will*. Oxford: Oxford University Press.
- Schlosser, M. (2015). Manipulation and the Zygote Argument: Another Reply. *Journal of Ethics*, 19, 73–84.
- Todd, P. (2013). Defending (a modified version of) the Zygote argument. *Philosophical Studies*, 164, 189–203.
- Todd, P. (2017). Manipulation arguments and the freedom to do otherwise. *Philosophy and Phenomenological Research*, 95, 395–407.
- Waller, R. (2014). The threat of effective intentions to moral responsibility in the Zygote argument. *Philosophia*, 42, 209–222.