The Replication Argument for Incompatibilism

In recent times, it has become increasingly common for incompatibilists about moral responsibility and determinism to invoke what have come to be called *manipulation* scenarios – roughly, scenarios in which a given agent's behaviour is somehow "set up" in advance by various powerful manipulators or designers working "behind the scenes". The proponent of a "manipulation" argument for incompatibilism puts forward the given scenario, hoping to elicit the judgment that the relevant agent cannot be fairly blamed for doing what he or she does. He or she then contends, typically, that there is no relevant difference between the agent in this (non-responsibility) scenario, and *any* agent in a deterministic world; there is no relevant difference between one's actions having been guaranteed in advance by powerful designers, or instead guaranteed by (otherwise similar) blind natural causes. Determinism thus precludes moral responsibility.

In my view, the basic insight behind the manipulation argument is important and compelling.[1] Many compatibilists, however, have not been convinced. Once one appreciates precisely the way in which such "designed" or "manipulated" agents may meet their preferred conditions on moral responsibility, they maintain, it will no longer seem clear that such agents could not be responsible for what they do. The incompatibilist judges that no such agent could be responsible; the compatibilist disagrees. Such is, broadly speaking, the state of the debate concerning manipulation arguments.

In this paper, I wish to move the debate forward, not simply by defending the ordinary incompatibilistic judgment in a standard manipulation case, but by building on the manipulation scenarios considered thus far to provide a new kind of argument for incompatibilism. As we'll see, though my argument is *inspired* by manipulation arguments, my argument is not a manipulation argument – or even an "original design" argument. In particular, my argument bypasses the most common objections to such arguments, while remaining just as powerful as those arguments – and arguably even more so.

In order to present my argument, I am going to tell a story – a story rather more elaborate than those usually told in these contexts. (The elaborateness of the story will help to address one of the most common complaints about manipulation arguments, viz., that it is unclear how we are to envisage the details of such cases.) The story is, I admit, fanciful – but it is no more (or barely more) fanciful than stories with which

[1] See Todd 2011, 2012, 2013, and 2017. See further, e.g., Pereboom 2014: 71 – 103.

anyone working in these debates is already familiar. The story is, in a sense, a version of the "rollback" scenario depicted by Peter van Inwagen.[2] In the rollback scenario, one is asked to imagine that God "rolls back" the state of the universe until just before an agent deliberates and makes a given decision. In this context, the assumption of determinism implies that, in every rollback, we get *precisely the same result*. The fundamental diagnostic element in the rollback scenario is therefore *time*: we continuously "reset" time to just prior to the agent's decision – and thus we see something about the nature of determinism. In my story, however, I wish to *spatialize* the rollback scenario. That is, instead of considering 10,000 (temporal) replays of the same set of initial conditions, I wish instead to paint a scenario in which one discovers 10,000 (spatially separated) distinct actualizations of these conditions. In considering this scenario, I believe that we come across a particularly powerful argument for incompatibilism. For reasons that will become clear shortly, I propose to call it *the replication argument*.

First, I tell my story. I then draw out in more detail the incompatibilist argument it makes possible, and respond to some objections. Next, I contrast the argument I develop with the "manipulation" and "original design" arguments already present in the literature. I conclude by suggesting how the present argument plausibly constitutes an advance over such arguments.

The Background

Alfred Mele has asked us to consider the story of a goddess who actualizes a set of initial conditions that (deterministically) guarantee that there comes to exist a certain person – Ernie – who subsequently performs various actions.[3] Now, once we do *that*, we can of course imagine such a goddess creating *many* distinct sets of these conditions, thereby creating many scenarios in which there come to be *distinct* "Ernies", all of whom live type-identical lives, down to the finest details. Instead of imagining a goddess with these sorts of powers, however, we might imagine that there is a "universe architect" that has arisen from a highly technologically advanced society of the very distant future. And, since we are philosophers, we can further imagine that all of these "universes" are somehow present on one physically located planet – that is, since we are philosophers,

[2] See van Inwagen 2000.

[3] Mele 2006: 188. In Mele's scenario, these "initial conditions" are not *comprehensive*; Diana is arranging conditions *within* the deterministic universe she inhabits. Of course, it is a small step from this case to a case of *total* creation, which is the kind of case I consider shortly. But this step may be important; I return to this point below.

we can imagine that the architect has a way of setting up what we might call "pods" on a given planet (not unlike, in some respects, the "pod" depicted in "The Truman Show"). Each of these pods represents, then, a deterministic universe unto itself – and each is suitable to house (and does come to house) a community of agents whose lives and activities, from the outside, appear to be exactly like our own. Now, once you've imagined the possibility of a planet hosting such pods, all of which represent deterministic universes with precisely identical initial conditions, you can further imagine *going* to that planet and *discovering* those pods. Indeed, a planet precisely such as this is the scene of the following story.

The Story

At long last, you've arrived. You've heard the rumors about Planet M – strange signals; signs of intelligent life. You're ready to explore. You get your landing vehicle ready and set out. After a while, you see something on the horizon – something that appears to be in the shape of a large dome. Your excitement builds: it would appear to be some kind of "pod". As you approach, you notice something curious: what would have seemed to be the "wall" of the pod is in fact strangely translucent. You can see through – and far inside, as you focus here or there. You dare not touch it. But soon your attention is drawn away, for you notice: there are people inside, people just like you. You stand transfixed. It dawns on you. The Civilization, or someone within it, must have generated a Universe Architect, precisely as you surmised. And this must be the Architect's creation; this must be the Architect's universe, a universe she *began* but which she has subsequently left to evolve on its own. And here you are – godlike – observing it. You notice that you can, if you wish, listen to the conversations of those inside.

You settle in, scanning for the big picture. Your initial impressions are, alas, rather grim. The inhabitants of this universe would seem to have made a mess of it. At the global level: wars, unaddressed poverty, systematic injustice, and yet more besides; at the personal level: backstabbings, bitter feuds, jealousies, lies, graft, and missed opportunities. Of course, it isn't all bad: there is also kindness, cooperation, and heroism. After a while, despairing of the big picture, you focus in more carefully, and soon you take an interest in a perfectly ordinary man named Robert, who is just beginning a new career. That career, you discover, commences under the auspices of a particularly unpleasant new boss named Barry. In fact, your first impression of Barry arises from

your seeing him deliberate concerning whether to play another round of golf, or instead make his meeting with Robert; he decides to play the golf, despite knowing that Robert will be waiting listlessly in the meantime – and when he finally arrives, he does so with a flood of lies about his car breaking down. Your temper flares. But it gets worse.

It turns out that Barry begins to funnel money into a secret account, ostensibly overseen by Robert. Barry is using him as cover to commit fraud – and his plan is airtight. If the authorities investigate, and they will, Robert will take the fall, and Barry will take the money.

It continues. The authorities make inquiries to Barry; Barry redirects them to Robert. Your blood begins to boil. The crucial moment is approaching. Unless something happens, Robert's life will be destroyed. You can't stand the thought. You've never thought to touch the prism, but now you can't resist. You look at the scene and thrust out your hand…

At that moment, from the point of view of pod-1367, a miracle occurs. You are transported to the scene. Securing the crucial files, you burst into the room and exclaim, "It was Barry!" Everyone stares at you, dumbstruck by the strange person that has suddenly appeared. Their shock triggers your alarm: what have I done? You dash back out of the room, and see the gap closing…

You are out. Once the adrenaline dies down, you're relieved to see Barry being led away in handcuffs…

After another week, you begin thinking about exploring more of Planet M. You pack up your vehicle and set off. You've barely gotten going when you see it. Your heart skips a beat. It is another pod.

Approaching, everything seems the same; the same translucent wall, the same strange prism. And once you look inside, you notice something uncanny. You've seen these scenes before. Indeed, you've seen *exactly* these scenes before; you are seeing precisely what you had already seen in the other pod – what you had thought was *the* pod. That war is going on over there in just the same way. The same poverty, the same injustice. And the same Barry. As you focus in, there he is, deliberating about whether to make the meeting with Robert. Not a molecule of the scene looks different. Here it comes. Yes, he just decided to skip the meeting… and here come the lies.

You think: what is going on? What is this place? You race back to the vehicle and keep going. Another pod! You investigate for 10 minutes; you see precisely identical scenes unfolding as you had previously. Your mind is spinning.

You see ahead of you what appears to be a hilltop; as you arrive at the summit, the scene unfolds before you: fields of pods, beyond the eye can see. There must be thousands of them. Tens of thousands. Calming down, you notice what seems to be a large monument bearing some sort of inscription:

The Story of This Place

It came to pass that there were the artists and the architects. The artist writes a story; the architect is employed to bring it to life in the form of a universe. The stories were at first simple, and the architects had various degrees of success in bringing them to life. But the architects got better. And as they did, the stories dreamt of by the artists became more elaborate. Fewer and fewer architects were able to ensure that the stories came to life as intended by the artists. Soon there was only one: I, Jane, author of this inscription. The most challenging story yet was presented. I brought it to life. But some said it was luck. Some said: do it again, and we'll believe that you are the Architect. And so I did it again. And again. What lies before any who might discover this place is a testament to my power to achieve what I intend. You will discover 10,000 universes unfolding before you, all precisely the same, all unfolding precisely according to my design. Do not suppose that it is by any form of chance that they are the same. For I have discovered the key. Be convinced of my power to determine.

- JANE (The Architect)

—

A few weeks later, you have begun to process the message on the monument. (You had a look at some more of the pods to verify the truth of the inscription.) Now you have occasion to think back to your discovery of your first pod. You are recalling your initial reactions. For a curious thing happened as you started going from pod to pod, once you became aware of the truth of the inscription. Your attitudes changed. In particular,

when you approached a particular pod and discovered Barry's duplicity and lies, you were still *sad*. Robert was still there, getting hurt, same as before. But something changed. It is just hard to say what it is.

You remember happening upon the first pod. When you looked inside, you eventually discovered all of the wars and the unaddressed poverty. And you remember thinking: these people have really made a mess of it. And you were, in that moment, subtly *condemning* them for making a mess of it. You had been thinking: this world is a mess, and it is your *fault* that it is. But after coming upon pod after pod after pod in which *precisely* these same wars were going on in *precisely* the same way, something happened. You stopped thinking that it was their *fault* that these wars were going on. Concerning the first pod, you might have said: I didn't have to discover this pod as containing all of these wars, but I *did,* and it is your *fault* that I did. The pod didn't have to turn out this badly, but it did. For this, you all are to be blamed. That, anyway, was your thought upon seeing the first pod – or, perhaps, it was underlying your thought. But now you wonder whether that reaction was really justified. For now you wonder whether you can really say: the pod didn't have to be this way, and it is your fault that it is. These wars didn't have to happen, and it is your fault that they have. Now it would seem that this judgment is called into question. When you come upon a new pod, and, sure enough, it contains just the relevant war at just the relevant time, you think: of course it does. It has to. The war comes to seem like a temporally extended natural disaster. Of course, it has its victims, and it has its instigators. You just can't bring yourself to condemn those instigators like you once did. Their instigation was written into the story.

And then there's Barry. You remember watching him deliberate concerning whether to keep playing the golf or instead make his meeting with Robert, as he should. When he blew off Robert, you resented him on Robert's behalf. But now you come up to yet another pod. You wait around. There's Barry, and there's Barry lying. But you no longer feel the same kind of *resentment* towards Barry. Of course, you think to yourself: Barry is still a jerk. He is still, descriptively, an awful person. You just can't *condemn* Barry for being a jerk, and for being as awful as he is. Of course he's that awful. Whose fault is that? Well, if it is anyone's, it is Jane's – or perhaps it is the artist's, who *wrote* Barry *as* awful. So, yes, of course he's awful and couldn't have failed to have been. Jane's display demonstrates that beyond reasonable doubt.

Still, you reflect. Suppose I intervened again into one of these pods, and suppose this time I got trapped. You think to yourself: well, I still wouldn't want to be Barry's *friend*. You feel like that's perfectly appropriate. You suppose that you wouldn't know how to interact with Barry. You would feel a sense of profound ambivalence, perhaps. Or perhaps at times you would simply forget what you knew about Barry: that he could not have turned out differently. From the *inside*, he doesn't *look* like someone who has to do what he does. Indeed, from the inside, there is no reason he "has to" do what he does: the artist's story doesn't involve Barry "having to" tell a lie. All the same. From the outside perspective, you know, there is a perfectly good sense in which he has to tell the lie. He has to with precisely the same force as the next pod *has to* perfectly resemble the former. And that knowledge dulls your resentment. When you remember Barry being led away in handcuffs, after your intervention, you remember being delighted that he's getting just what he deserves. Now you wonder. Of course, nothing you've learned implies that what Barry was doing was in fact fair, or in fact was never going to harm Robert. But still you wonder. What does it mean to say, now, that Barry got what he deserved? Internal to the *story*, of course, Barry deserved to get caught; internal to the story, he had no excuse. At some deeper level, however, you think: Barry does not deserve anything at all. Hence your ambivalence about his now, thanks to you, being in jail.

Reflections

Well, there is my story. But I promised an argument for incompatibilism. My argument, I suppose, is this. Your reactions in the above story are the right ones. Those reactions, however, are incompatibilist. Naturally, this argument will strike you as strong only to the extent that you sympathize with the reactions I presented. What might be said further on their behalf?

The central intuition my story is meant to bring out is the following. Though each of the pods *is* a given way, no one *inside* those pods is at fault for its *being* or *turning out* that way. Pod-1367 turns out to contain a lie from a person named Barry to a person named Robert. Now, whose fault is that? Ordinarily, we might suppose: it is Barry's fault that it turns out that way. Once we appreciate the wider context, however, it no longer seems correct to say that it is Barry's fault that pod-1367 contains a lie from a person named Barry to a person named Robert. It does contain such a lie, but it is not

Barry's fault that it does: this is Jane's fault, if it is anyone's fault. But if it is not Barry's fault that his universe is characterized by a lie from him, he is not fairly blamed for the fact that it *does* contain such a lie. And this seems to be another way of saying that he is not morally responsible for lying – or, indeed, for anything at all.

Perhaps the best way to bring out this judgment (insofar as anyone may be brought to have it that doesn't already) is the following. Suppose we modify the story. Suppose you arrive at M and discover the following monument:

> *The Story of This Place*
>
> It came to pass that there were the artists and the architects. The artist writes a story; the architect is employed to bring it to life in the form of a universe. The stories were at first simple, involving only inanimate objects and their interactions over time, and the architects had various degrees of success in bringing them to life. But the architects got better. Eventually, they could reliably produce any such story dreamt of by an artist. But the stories dreamt of by the artists became more elaborate. Indeed, artists now wished their stories to include *agents* – people. But, if it had agents, no architect could reliably produce exactly the story the artist wished; if a universe with agents unfolded just as desired, that was, for the architect, mere luck. It couldn't be reproduced at will. I, Jane, have come as close to reliability in this matter as anyone could come. What lies before any who might discover this place is the following: all universes with precisely the same initial conditions, all guaranteed to produce agents. But I cannot say precisely what those agents shall decide to do or to become. As you investigate these pods, that is what you will discover. Each has begun from the same seed. But see, now, what they have made of themselves and the universe I designed.
>
> - JANE

Now, against *this* background, I submit that our attitudes towards those we discover in the "pods" would (or should) be different (as compared to our attitudes in the first story). If you discover a war in one pod, but not in another, then you can continue supposing that it is their *fault* that their pod contains a war. It didn't have to. Other things being equal, we could look at the other pod and say: they should have made

themselves like *that*. Nothing *yet* tends to engender in us any ambivalence about how we ought to regard those in the relevant pods. We simply assume, as you did in your initial encounter with the pod above: they are responsible for how their pods turn out. Of course, a philosopher may come along and present to us what is commonly known as the "luck objection" – and at that point, we may begin supposing that this new story is no better than the old one as regards how ambivalent (or not) we ought to be.[4] But this is another matter.

Perhaps I can put my central point as follows. In the original story, we arguably become (or ought to become) ambivalent – and our ambivalence arises from the following tension. On the one hand, we recognize that there is no *story-internal* reason why Barry "has to" tell the lie. He doesn't feel like he has to. No one is forcing him to tell the lie. Indeed, we might imagine that the "artist" who gave the script to Jane wrote it (at least in part) as follows: "… and then Barry tells the lie, because he wanted to stay out playing golf, even though he didn't have to." And so Jane brings about exactly that. Such a universe looks much different than a universe in which the script says: "…and then Barry tells a lie, because of his psychological condition, and so he has to."[5] On the other hand, there is a reason *external* to the story that Barry *has to* lie, and herein lies the tension. In light of our further knowledge, we can no longer judge that Barry *should have* made his universe a non-lying universe: that is, we now think, quite beyond Barry's power, exactly inasmuch as it is *within* Jane's power to get precisely what she wants. Barry's power to make his universe a non-lying universe is the mirror opposite (or the inverse) of Jane's power to secure what she wills. And that power, we recognize, is *infallible*. From a standpoint internal to Barry's universe, it appears that he has no excuse. Looking from the outside, however, he is excused.

But precisely the force of the modified story is this. If there are other universes precisely like Barry's (up to the relevant time) but in which his counterpart tells the truth, then, other things being equal, he is not excused even from the external perspective. We can, even from the outside, still regard it as his *fault* that his has turned out to be a

---

[4] The "luck objection" to incompatibilist theories of responsibility contends that indeterminism simply implies *randomness*, and therefore cannot help with freedom or responsibility. Accordingly, indeterministic universes (however described) can allow for no more responsibility than deterministic universes. For a recent response to the luck objection, see Franklin 2011.

[5] Further, we may certainly grant that *if* Barry didn't want (or try, or…) to lie, he wouldn't lie. It isn't as if Jane or the artist (or something or someone else) is around to stop him from telling the truth, *if* that is what he wanted to do (or had tried to do). Indeed, it is precisely Jane's genius that she can, in this way, fix what shall happen in each of the pods, without needing to sit around and *monitor* or *interfere* with those pods.

universe in which he lies. There is no story-internal reason why Barry had to lie – and no story-external reason either. The full range of our moral-sentimental reactions and judgments, however, would seem to presuppose that the relevant agents are not excused from *either* perspective. The full range of these attitudes and judgments, therefore, is plausibly incompatible with determinism.

Five Objections

Here is the first:

> You just concluded that the full range of the reactive attitudes is incompatible with determinism – determinism *as is,* or in itself. Strictly speaking, however, what you've shown – if you've shown anything – is that the agents in these deterministic "pod-universes" aren't responsible. You aren't yet entitled to conclude anything about *determinism itself,* or other deterministic universes that are *not* designed or produced in the relevant way (such as, perhaps, our own). Evidently, in order to conclude something about these other deterministic universes, you need some sort of bridge principle, a principle that says that there is no morally relevant difference between the given pod-universes, and other deterministic universes that are simply "there" as a matter of brute natural chance (or anyway not as a result of some kind of cosmic design). In other words, you need it that whether or not a deterministic universe was the product of the relevant kind of design is not in itself relevant to the moral responsibility of the agents inside those universes. But perhaps it is relevant in just this way. Perhaps you could have a qualitative duplicate of Barry-1367 in a different (yet similarly deterministic) universe, and that other Barry is responsible for what he has done, and criticisable, whereas Barry-1367 is not, precisely because Barry-1367 inhabits a universe that was designed, and the other Barry does not.

The bridge principle is needed, but the bridge principle seems plausible. Consider the final claim of this objection. Doesn't this claim seem suspicious? Suppose you had mixed up which of the two universes was which. Now imagine intervening into the latter universe, and upbraiding Barry for what he has done to Robert. After all, he has met all of the relevant compatibilist conditions for responsibility, and it isn't as if his

universe was *designed*. But then the news comes in: actually, this *is* the designed universe – and so one abruptly apologizes to Barry ("Just kidding!"), for one's new information implies that he is not responsible for what he has done after all. Such a shift seems implausible. And its implausibility reveals the following general thesis: responsibility is suitably *intrinsic* – and its being suitably intrinsic implies that facts wholly *extrinsic* to the causal processes that constitute an agent's total life are not themselves relevant to whether that agent is responsible for any actions taken within that life. And, ex hypothesi, whether the universe was produced by design or not is extrinsic to the causal processes in the universe itself. (This is why the given universes are *intrinsic* duplicates.) Of course, much more might be said about these issues[6], but I take it that most readers – compatibilist and incompatibilist alike – will grant that the first objection can be answered.[7] On compatibilist assumptions, why should it matter that Barry's universe was produced in the way it was?

Here is the second objection:

> You contend that, on discovering the rest of the relevant (identical) pods, our indignation would (or should) dissipate into something like moral sadness. But imagine going into one of the pods and attempting to explain to *Robert* that his indignation with Barry is misplaced. Imagine trying to tell *him* that he is making a mistake. Reflection on what we can tell Robert, in this way, indicates that his indignation is *not* a mistake. So: neither is ours, even if determinism is true.

But this objection confuses what the incompatibilist wishes to show. Needless to say, the incompatibilist does not judge that Robert is himself *blameworthy* – or otherwise criticisable – for blaming Barry. After all, Robert's blaming Barry, in precisely this way, is yet one more part of the artist's script. The idea is not that Robert is doing something which he is criticisable for doing, which would license us charging him with "making a mistake". The idea is that, unbeknownst to Robert, the person he is (blamelessly) blaming does not, in fact, deserve this blame – *even if* this is a fact that Robert himself could not possibly be brought to know. In short: we cannot confuse someone's blame

[6] As an anonymous referee for this journal has noted, there is a sense in which it is intrinsic to "the causal process" – totally conceived – the led to Barry-1367's action that it traces back to the casual activity of Jane. However, when we focus on the causal processes that constitute Barry's life, Jane's activity is extrinsic to these processes. And this, I contend, is what matters.

[7] I return briefly to similar issues below, when contrasting my argument with Pereboom's "four case" argument and Mele's "Zygote Argument."

being understandable, or uncriticisable (and so in that sense "appropriate"), with that blame being *deserved* by the person blamed (and so in *that* sense appropriate). One might be (rightly) reluctant to tell someone recently abused that his abuser was a mentally unstable psychopath, and so is not to be blamed – while still feeling that, in the final analysis, that person *isn't* to be blamed.

Consider. Suppose one could bring Robert out of pod-1367 – and suppose one could show Robert the other pods and the relevant inscription. *Modulo* the profound disorientation such a revelation would have to produce (I return to this theme shortly), we might consider what sort of reactions it would be appropriate for Robert to have. Would it be appropriate for Robert to blame Barry-1368 (in the next pod over) for what he has done to Robert-1368? It seems to me that the answer is plainly "no" – and the argument for this claim mirrors the one already provided: on coming to see the other duplicate pods and on understanding how they were produced, Robert should judge that the given Barry could not have turned out differently, or, more particularly, that it is not his *fault* that the given pod is characterized by a lie from him to a person named Robert. But when Robert makes this judgment, he should go on similarly to judge that *his* Barry – the Barry who lied to *him* – was no more responsible for doing so than the others are (which is to say: not at all). Once more, such a judgment would be profoundly disorienting. But then, well, determinism has profoundly disorienting consequences – or so it seems to me. It is precisely these consequences I am trying to bring out.[8]

Here is the third objection:

> In the actual world, people often experience "moral desensitization", but the best explanation of this fact is not that people are eventually judging that the relevant parties simply aren't responsible. We have finite emotional resources, and it can be exhausting to remain staunchly indignant, for instance, at political scandal

[8] It is worth remarking on here precisely what would be the source of the "disorientation" at issue. Consider the "libertarian" variant of the scenario as imagined above. Now, imagine that one is somehow brought out of one's pod and made to see the *other* pods – but that these other pods simply contain different events and different people. Such a revelation would, perhaps, be deeply strange – I've been living in a pod this whole time! And there are all of these other pods with different things going on! And all made by the same super-intelligence! – but it seemingly would not be profoundly *disorienting*, and this is because, so far, this discovery does not in any way suggest that one had been subject to any kind of illusion or that one's self-conception was somehow mistaken. What is so powerfully disorienting about Robert's experience (in the deterministic scenario) is precisely his realization that all of the people he had been dealing with were, in a sense, robots – and, by extension, that *he* was (is still?) such a robot. In other words: it is the revelation that everything he and everyone else had done had been determined that is disorienting.

> after political scandal. As Strawson noted, sometimes we may retreat to the "objective stance" as a "refuge from the strains of involvement."[9] But this is, plausibly, what is going on in this story: on encountering pod after pod after pod, of course one's anger eventually dissipates, and perhaps becomes mere sadness. But what would explain this would simply be one's moral exhaustion in keeping up the "reactive stance" in the face of so much criticisable wrongdoing, *not* one's judgment that the relevant agents simply aren't responsible in the way one had initially thought.

But this is not, I contend, a plausible explanation of the facts of my story. Note: in my story, one first discovers pod-1367, and becomes indignant with Barry. One then discovers another (identical) pod. One then discovers the *inscription* and the whole truth about all of the pods. (There is no reason to think this process must take a great deal of time, nor a great deal of emotional energy.) Now, my contention is that it is *this discovery* – the discovery about Jane and the nature of the pods – that in itself dulls one's indignation. As one approaches the next pod, one approaches it, not emotionally exhausted, but instead attuned precisely to the way in which Barry could not have turned out differently. It is this that makes one ambivalent.

Of course, my story is just one story. We could invent others. For instance, we could imagine a story in which a latent compatibilist travels to Planet M and makes the relevant discovery. We could imagine this person discovering pod-1367, becoming indignant, and then discovering the relevant inscription, and continuing on in her indignation towards each subsequent Barry she discovers – anyway for a time, until her emotional energy runs out. Note: here we are not to imagine this person, as she travels from pod to pod, remaining angry *that this is happening* – that, for instance, there is an innocent person named Robert who is suffering unjustly. No. We must imagine this person remaining angry *with each Barry* that this is happening. We must imagine her continuing to feel like it is *his fault* that this is happening. It is, of course, easy to see how one might persist in moral anger – of *some* description – on discovering pod after pod. It is, I think, considerably harder to imagine persisting in moral anger *with the relevant Barrys* that it is happening. But this is the stance to which the compatibilist is committed. To what extent this is a cost for compatibilism the reader must decide. More generally: the

[9] For discussion of this Strawsonian theme, see Tognazzini 2015.

reader must decide whether he or she sympathizes more with the reactions of the person in the story as I presented it, or with the reactions we have just considered.

But here we come to the fourth objection:

> I do not want to be a through and through skeptic about "thought experiments" in philosophy. But here you're asking us to consider what our reactions should be under a scenario in which we somehow peer inside 10,000 carbon-copy deterministic universes – and, just a moment ago, you were implicitly inviting us to consider what we should think if *we* were taken "outside" *our own* universe and shown such carbon copies (of ourselves) ourselves. But at this stage, we have arguably reached the end of our imaginative capacities; and so who can possibly say what our reactions would or should be in these bizarre scenarios? Perhaps vertigo[10] is the only proper "reaction"; it is certainly the only predictable reaction.

The first thing to say in response to this objection is simply to downplay the bizarreness or inconceivability of the stories I have imagined. You have seen a space exploration movie? Good; so you can imagine going to a far away planet. You have seen *The Truman Show*? Good; so you can imagine peering inside a (relatively self-contained) "world" and watching its goings-on. You can imagine becoming indignant at what those people are doing? Good too. Now, from this, you can imagine going a "pod" over and discovering a total duplicate of what you had just seen. And can't you imagine accepting that this result is the deterministic consequence of the designs of some higher intelligence? This is, admittedly, more difficult, but I will accuse you of a suspiciously selective lack of imaginative abilities if you claim that you can't. But this is all that is required for the setup of my story. I don't mean simply to ignore the complexities that would have to be involved to make the story a fully detailed narrative. However, I do claim that it is as coherent and intelligible as other "thought experiments" we often entertain in similar ways; I am confident, for instance, that a highly skilled film team could create a believable (if terrible) Hollywood film that takes my story as its key premise. We often morally evaluate the actions and reactions of characters in such fictions; why, then, not my own?

---

[10] Cf. Watson's (1999: 364) apt remark: "To think of oneself as at once a full fledged free agent and as a creature whose every move, every hope and scheme is part of another's plan is certainly vertiginous."

But my more fundamental response to such a complaint is this. Whether or not the complaint is legitimate, it is not a complaint anyone who wishes to have an opinion on the question of the compatibility of responsibility and determinism has a right to make. More particularly, if you ask whether something familiar is compatible with something bizarre, then you can't complain when the resolution of this question requires thinking about something bizarre. And the doctrine of determinism is, in this sense, bizarre – not in the sense of implausible or unmotivated, but *hard to imagine* and *hard to comprehend* and *epistemically distant*. According to the doctrine of determinism, your deciding not to read this paper would have required a change in the past – indeed, a change *all the way back* to the big bang. The very initial conditions of the entire universe would have had to have been different, for you not to read this paper. (Or, on a different approach, some *law of nature* would have had to have not been a law.) This is hard to comprehend. This is – in the relevant sense – bizarre. If you want to know whether responsibility is compatible with the truth of something like *that*, then, once more, you can't complain when the resolution of this question requires thinking through something bizarre. We are *already* in the realm of the bizarre.

Consequently, I wish to resist the suggestion that the strangeness of the scenario somehow favours the compatibilist, or otherwise permits her simply not to consider it. If you say, "I have no idea whether I would continue to hold Barry-1367 responsible for what he did to Robert-1367, on coming to see the inscription and the rest of the carbon-copy pods", then you have no idea, at this moment, whether, deep down, you are a compatibilist. Or if you say, "I have no idea whether I *should* continue to hold Barry-1367 responsible for what he did to Robert-1367, on coming to see the rest of the carbon-copy pods", then you have no idea, at this moment, whether compatibilism is true. Indeed, if this is your reaction, then this is strong evidence that the argument I have presented is a good one, at least by your lights, since it has made you, at least for the moment, agnostic about the truth of compatibilism.

Here, then, is the fifth objection:

> It is always possible to "reverse" an argument – to say modus tollens where another says modus ponens. And, in this case, once we see how the compatibilist ought to "reverse" the argument, we'll see that, in the end, it doesn't bring out anything more counterintuitive about compatibilism than what we had implicitly already been accepting. Consider. Suppose we feel sure that *we*

> are responsible. Now suppose we are somehow made to know that there are 10,000 other universes just like our own. Now, you want us to say, "Since there are those other 10,000 carbon-copies of us, we aren't responsible." But couldn't we with equal justice say, "Because *I'm* responsible, so are those other 10,000 carbon-copies?" After all, if I feel responsible for doing what I've done in the circumstances I've faced for the reasons I've recognized, why shouldn't anyone *else* who has also done the same thing in those circumstances for those reasons? In short: I'm responsible, that person is a carbon-copy of me, and so is also responsible. That's better than: That person is a carbon-copy of me, and isn't responsible, and so neither am I.[11]

But this "objection" simply misrepresents the epistemology of the argument as I have presented it (or anyway as I intend it). Note that I did not – and for precisely this kind of reason – start with a scenario in which one discovers copies of *oneself* and one's own universe, and consider what one should say in that kind of case. (I considered this sort of thought only later, and in response to an objection; if you haven't liked what I've said since then, then I invite you simply to forget it.) Such a procedure immediately introduces a kind of bias into the evaluation of the responsibility of the relevant agents, for in evaluating *their* responsibility, one takes oneself also to be evaluating one's own – and it is difficult (if not impossible) to evaluate oneself as non-responsible. This is why I have told the story as one in which one discovers a given universe of which one isn't a member, evaluates an agent in *that* universe as morally responsible, and then discovers the carbon-copies of *that* universe. One then must consider: do I go back on my initial judgment that the agent I first encountered was to blame? Or do I persist in this judgment – and maintain that all the given agents are to blame? And I have tried to bring out the plausibility and attractiveness of the former judgment. One's own responsibility, at this stage, simply isn't in view. This is, indeed, as I see it, one of the core devices behind the argumentative strategy I have in mind.

In this light, the more accurate way to attempt to "reverse" the argument I have in mind is this:

> One first evaluated Barry-1367 as responsible. One then discovered the other pods and the other (carbon copy) Barrys. Now, you want to say that those other

[11] This objection is inspired by (but only very loosely modelled on) replies to the "manipulation argument" offered by McKenna 2008 and Fischer 2011.

> Barrys show something about Barry-1367, but why can't one say that Barry-1367 shows something about them – namely, that they're just as responsible as one initially took *him* to be?

But now we are simply back to our previous question: do you, or do you not, sympathize with the reactions of the character in the initial story I presented? Needless to say, if you don't, then if an argument for incompatibilism will move you, it will not be this one. But the point is this. The idea is that the other pods are meant to be *revelatory* in some way – that they reveal a truth about Barry-1367 to which one had previously not been sensitive. If you think they reveal something, but what they reveal is something trivial, then, again, if an argument will move you, it is not this one. But it may yet move others – in particular, those who haven't yet considered the issue of responsibility and determinism.

A final note is this. There is nothing in the thought itself of there being other universes precisely like one's own that tends to call one's responsibility into question. It is, rather, what *explains* why there are (or would be) such universes that tends to do so. Consider. Suppose one thought that one had "libertarian free will", and suppose one thought one was responsible for what one has done with the libertarian free will one has. Well, why should it matter if someone *else* with libertarian free will, in a different universe, *also* has done precisely the things one has done? Obviously, it wouldn't matter – at most, again, it would be (Twilight-zone) unsettling. In my story, what matters is not just that one discovers a new pod in which – lo and behold – we have all the same things, but that one discovers the *explanation* of this fact, viz., that they *had to* be the same. It isn't just that there happen to be these other universes, it is that the given universes are the products of a common cause, in whose power it was to *make* them the same. Look at it this way. If one is watching Jane construct yet one further pod – pod 10,001 – will one judge that it is going to be the given Barry's fault that his turns out to be a universe in which he lies to a person named Robert? I contend that this judgment will seem forced in this context. And what will make it forced is this: one knows that his universe will be a copy of the others, and one knows the deeper explanation concerning what guarantees that this is so.

Advantages and Advances

At this stage, I turn briefly to comparing the argument I have developed above with the so-called "manipulation" and "original design" arguments which have inspired it. Here I consider two such arguments, with which I assume broad familiarity: the four-case "manipulation" argument of Derk Pereboom, and the "original design" Zygote Argument of Alfred Mele.

By way of review: Pereboom's argument asks us to consider a series of cases in which a "team of neuroscientists" manipulate a certain agent (Plum) into performing a certain action (killing White), all the while, the argument contends, meeting compatibilist conditions for responsibility. In Case 1, the neuroscientists control Plum by means of "radio-like technology" and a chip in his brain; in Case 2, they achieve a similar effect, except that they have "pre-programmed" Plum at birth to be "rationally-egoistic" in the requisite way. In Case 3 (which, for various reasons, seems less central to Pereboom's case, and accordingly gets less frequent attention), there are no neuroscientists, but nevertheless a similar effect is achieved via Plum's atypical upbringing. And Case 4 is simply a "normal" deterministic case in which Plum kills White in the same sort of way as he does in the other cases, but without any "manipulation" in the background. The idea is this: insofar as one judges that Plum isn't responsible in Case 1 (as one is meant to), this judgment gives one reason to think Plum isn't responsible in Case 4, for there are (meant to be) no morally relevant differences between the cases. But since Case 4 is simply the normal deterministic case, if Plum isn't responsible in that case, compatibilism must be false.

The first thing to say about my argument as compared to Pereboom's is simple. My argument features no "manipulation", and is thus not a manipulation argument. Jane does not, in any sense, "manipulate" Barry into lying to Robert: she simply creates the (deterministic) circumstances in which he *will* lie to Robert. But Jane does not interfere with Barry or uniquely "pre-program" him or anything of the sort. (Of course, one may stipulate a sense of "manipulation" in which Jane "manipulates" everyone in the relevant pods, but this sense of "manipulation" seems attenuated at best. Similarly, though there is a sense in which Barry's *whole universe* is "pre-programmed", it does not follow that Barry himself, and his actions, are uniquely pre-programmed *within* that universe.) Accordingly, my argument simply bypasses the most common worry about "manipulation" arguments, viz., that the envisaged manipulation in fact does not leave the agent's compatibilist-friendly capacities intact. These worries are well-known in Case

1.[12] But they might also reasonably arise even in Case 2 (which is the case that bears most resemblance to my own). Consider. Arguably, the "non-responsibility" intuition is most apt to be generated if the result *that Plum kills White* is itself within the neuroscientist's power to guarantee: having played their part, they can simply sit back and relax. (Notably, Pereboom compares his Case 2 to the case of a Leibnizian God.[13]) So suppose they "program" him (in some such way) *now*. But now they ask: well, what will happen if, in the intervening years, Plum is, say, kidnapped and taken to a religious commune? How can we guarantee that Plum will return to kill White *then*? Arguably, only by making sure that Plum will be *extremely* psychologically abnormal. At a more mundane level: what if Plum is simply killed in the meantime – hit by a bus on the way to work?

Reflection of this kind leads one to realize that, in Case 2, the neuroscientists in fact will have to be something close to omniscient in order to achieve their plan – but at this point, the case becomes strained, and we arguably transition to the sort of case designed precisely in order to evade this kind of worry (as well as the worries about "manipulation" noted above): an "original design" argument in the style of Alfred Mele's "Zygote Argument" scenario. Mele sets up his scenario as follows:

> Diana [a goddess with special powers] creates a zygote Z in Mary. She combines Z's atoms as she does because she wants a certain event E to occur thirty years later. From her knowledge of the state of the universe just prior to her creating Z and the laws of nature of her deterministic universe, she deduces that a zygote with precisely Z's constitution located in Mary will develop into an ideally self-controlled agent who, in thirty years, will judge, on the basis of rational deliberation, that it is best to A and will A on the basis of that judgment, thereby bringing about E (Mele 2006: 188).

But note the following. On Mele's construal, Diana is yet one more "thing" in the deterministic universe she inhabits. She is located in a particular place and particular time in that universe. How is it, then, that she is able to know the *complete state* of that universe at the given time – which is what she would need to know in order to be able confidently to deduce the relevant outcome? And in deducing *this* outcome, isn't she

---

[12] See Demetriou 2010. Pereboom has modified his case in reply to such worries; see his 2014: 76 - 77. See also Matheson 2016.
[13] 2014: 77

thereby deducing something about *her own* future actions? But here we encounter various paradoxes of agency. Arguably, we simply aren't considering this sort of worry because we are implicitly considering the case as a case of *total creation* (if Diana is fixing *this,* she'd have to also be fixing everything else), relevantly similar to the cases I have described above. But then, plausibly, we ought simply to *explicitly* consider such cases.

I'll cut to the chase. In my estimation, the cases we should be considering should either be Case-1 style cases of direct, "hands-on" interference and control, or they should be cases of *total creation*. Case 2 and the "Diana Scenario" are, however, problematically intermediate between these two kinds of cases, and are thus subject to various difficulties. More particularly, we have two paradigms of "control" to consider: God-in-relation-to-Pharaoh[14], and God-as-total-creator. In the former case, God is directly interfering with Pharaoh by "hardening his heart" – and it might reasonably be supposed that God could do so while respecting various compatibilist conditions. This, then, is the impetus behind the "manipulation argument" (though the trick for *this* argument will be getting all the way to incompatibilism). But in the latter sort of case, God is simply *creating* the initial conditions sufficient to determine the entire course of the universe. This is the paradigm at issue in my story. Case 2 and the Zygote Argument scenario approximate this paradigm, but, in my judgment, not closely enough.

But now we might note a further difference between my argument and Mele's. In Mele's story, Diana creates the relevant zygote *so that Ernie will perform an action at a particular time*. In my story, however, there is no reason to suppose that Jane creates the pod *so that Barry will lie to Robert*. Instead, she creates a pod *in which* Barry will lie to Robert. This fact may have been utterly inconsequential to the artist who wrote the script, and to Jane (who brought it to life). Indeed, perhaps she didn't even notice it. Now, I do not know how to settle the question of what makes an argument an "original design" argument. But certainly, in the literature so far, the focus has been on scenarios in which the given action under investigation is somehow *uniquely* fore-ordained or intended – and the discussion has focused on what intuition these facts bring out or ought to bring out.[15] Though my argument could incorporate such considerations, it

[14] Cf. Pereboom's (2014: 95) excellent reflections on a certain hard-line Calvinist preacher's position on this score. A further note: we might ask what attitudes it would be appropriate for *God* to have towards God's creatures in these scenarios – and we might further ask what attitudes it would be appropriate for *Jane* to have towards the relevant Barrys. After all, according to the compatibilist, Jane has made them both free and responsible. For an incompatiblist argument moving from similar themes, see Todd 2012, and for a reply, see McKenna forthcoming; see also Russell 2010 and Todd 2018.

[15] Cf. Fischer 2011 and Todd 2013.

does not depend on them; it does not depend on the fact that Barry, just now, is doing what anyone wanted to happen or intended to happen, and it does not focus on what intuition *these* kinds of facts bring out. Instead, it focuses on what the given *replications* bring out. To this extent, my argument importantly differs from the argumentative strategy incompatibilists have been pursuing thus far. Further, some critics of the Zygote Argument have contended that this very fact – that Ernie is doing just what Diana "effectively intended" him to do – constitutes a responsibility-relevant difference between Ernie and an otherwise similar agent who was determined by merely natural causes.[16] Now, I don't here mean to endorse the merits of this criticism of the Zygote Argument (in fact, as my discussion of the first objection would indicate, I think it fails). I mean only to point out that, whatever its merits, the current argument bypasses it. If such critics of the Zygote Argument wish to criticize the replication argument, they will have to press elsewhere.[17]

Conclusion

At this stage, however, one might ask the following. Suppose we modify the Zygote Argument scenario so that it is indeed a case of total creation. And suppose we consider an action of a person in the relevant universe that was, for Diana, inconsequential. So far, we have something similar to what has been considered before. Now, what is "added" by adding in the relevant duplicates, as in my story?

In my opinion – and with these reflections I conclude – the story I have imagined does not constitute simply a minor modification to the Zygote Argument scenario. As I see it, the story has the power to bring out in a compelling way precisely the sense in which many incompatibilists have thought that compatibilism is "morally shallow" – that determinism simply implies "the mere unfolding of the given".[18] Of course, the challenge for any incompatibilist making such claims is to give some content to the feeling of "mere-ness", so to speak – that what we have, under determinism, is *mere* unfolding – where this explanation is not simply a way of repeating the claim that

[16] See Waller 2014. For a similar objection to the Zygote Argument, see Schlosser 2015.
[17] One criticism of the Zygote Argument that would also apply to my argument can be found in Barnes 2015. In short, Barnes would contend that, in my story, Barry is not "potentially creative" (because, roughly, everything he does has previously been thought of by Jane), and being "potentially creative" is necessary for responsibility. My reply is this: either being "potentially creative" is not necessary for moral responsibility, or, if it is, then from the compatibilist perspective, it would be *ad hoc* to maintain that Barry is not "potentially creative".
[18] Smilansky 2000: 52 – 55.

determined events are, well, determined. But the story may help to give content to this feeling. On considering this story, we see that compatibilism presents us with a picture on which moral responsibility is wholly *reproducible*, and therefore comes on the metaphysical cheap. That is, we get a picture on which moral responsibility – indeed, blameworthiness – can simply be *generated*, and generated at will. It simply rolls off the metaphysical assembly line. If Jane wants a morally responsible Barry who is blameworthy for telling a lie to a person named Robert, that's what she gets. If she wants another one, that's what she gets. And another. And another. And so on, ad infinitum. She is, in a sense, simply *generating* blameworthiness. But there is some feeling – however inchoate – that moral responsibility should not have this kind of structure, or should not be reproducible in this sort of way. There is some feeling that, if it exists at all, it is a *deep* feature of reality – deep in such a way that would imply that it cannot simply be gotten so easily. According to the incompatibilist, moral responsibility goes deeper than this. If Jane wants a Barry who is morally responsible for lying to a person named Robert, there is no mechanical formula that will give her one: she must simply do her part, and then *wait*. If responsibility arises from the deep, then it does, but if it doesn't, it doesn't. At any rate, it is not simply hers to summon. As the story brings out, responsibility is not, in this way, reproducible. Compatibilism says otherwise. Compatibilism is false.[19]

References

Barnes, Eric. 2015. "Freedom, Creativity, and Manipulation," *Noûs* 49: 560 – 588.

Demetriou, Kristen. 2010. "The Soft-Line Solution to Pereboom's Four-Case Argument," *Australasian Journal of Philosophy* 88 (4):595-617.

Fischer, John Martin. 2011. "The Zygote Argument Remixed," *Analysis* 71: 267 – 72.

Franklin, Christopher Evan. 2011. "Farewell to the luck (and Mind) argument," *Philosophical Studies* 156: 199 – 230.

Matheson, Benjamin. 2016. "In Defence of the Four-Case Argument," *Philosophical Studies*

[19] For comments on previous drafts of this paper and/or helpful discussion, I wish to thank Neal Tognazzini, Philip Swenson, Garrett Pendergraft, Zac Bachman, John Martin Fischer, Justin Capes, Mark Balaguer, Yishai Cohen, Alfred Mele, Alex Pruss, Kenneth Boyce, Brian Rabern, and Justin Coates, with whom I first discussed the story told above on a hike to the 'C' in Riverside's Box Springs Mountains in 2008. Thanks also to participants at reading groups at Rutgers (including Eli Shupe, Dean Zimmerman, Holly Smith, David Black, Brian Cutter, Daniel Rubio, and Pamela Robinson) and Fordham (especially Joe Vukov and Amy Seymour), and members of the Ethics reading group at the University of Edinburgh, including Guy Fletcher, Debbie Roberts, Mike Ridge, Matthew Chrisman, and Elinor Mason.

173: 1963 – 1982.

McKenna, Michael. 2008. "A hard-line reply to Pereboom's four-case manipulation argument," *Philosophy and Phenomenological Research* 77: 142 – 159.

McKenna, Michael. Forthcoming. "Resisting Todd's Moral Standing Zygote Argument," *The Philosophical Quarterly*.

Mele, Alfred. 2006. *Free Will and Luck*. Oxford: Oxford University Press.

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

Russell, Paul. 2010. "Selective Hard Compatibilism." In *Action, Ethics, and Responsibility*, edited by Joseph Keim Campbell, Michael O'Rourke, and Harry S. Silverstein (MIT Press).

Schlosser, Markus E. 2015. "Manipulation and the Zygote Argument: Another Reply," *Journal of Ethics* 19: 73 – 84.

Smilansky, Saul. *Free Will and Illusion*. Oxford: Oxford University Press.

Todd, Patrick. 2011. "A New Approach to Manipulation Arguments," *Philosophical Studies* 152: 127 – 133.

Todd, Patrick. 2012. "Manipulation and Moral Standing: An Argument for Incompatibilism," *Philosophers' Imprint* 12: 1 – 18.

Todd, Patrick. 2013. "Defending (a modified version of) the Zygote Argument," *Philosophical Studies* 164: 189 – 203.

Todd, Patrick. 2017. "Manipulation Arguments and the Freedom to do Otherwise," *Philosophy and Phenomenological Research* 95: 395 – 407.

Todd, Patrick. 2018. "Does God have the moral standing to blame?", *Faith and Philosophy* 35: 33 – 55.

Tognazzini, Neal. 2015. "The Strains of Involvement," in Clarke, McKenna, and Smith, eds., *The Nature of Moral Responsibility*. Oxford: Oxford University Press.

van Inwagen, Peter. 2000. "Free Will Remains a Mystery," *Philosophical Perspectives* 14: 1 – 20.

Waller, Robyn. 2013. "The Threat of Effective Intentions to Moral Responsibility in the Zygote Argument," *Philosophia* 42: 209 – 222.

Watson, Gary. 1999. "Soft Libertarianism and Hard Compatibilism," *The Journal of Ethics* 3: 351 – 365.