



Manipulators and Moral Standing

Benjamin Matheson¹

Received: 20 June 2018 / Revised: 12 August 2018 / Accepted: 18 September 2018
Published online: 03 October 2018
© The Author(s) 2018

Abstract

Manipulation arguments aim to show that compatibilism is false. Usually, they aim to undermine compatibilism by first eliciting the intuition that a manipulated agent is not morally responsible. Todd's (*Philosophers' Imprint*, 12(7): 1–18, 2012) Moral Standing Manipulation Argument instead aims to first elicit the intuition that a manipulator cannot blame her victim. Todd then argues that the best explanation for why a manipulator cannot blame her victim is that incompatibilism is true. In this paper, I present three lines of defence against this argument for those who agree a manipulator cannot blame her victim.

Keywords Manipulation · Moral standing · Compatibilism · Incompatibilism · Moral responsibility · Blame · Blameworthiness

Many find it intuitive that a person who is manipulated to perform an action A is not morally responsible for A-ing. Incompatibilists about moral responsibility and causal determinism have capitalised on this intuitive thought in their attempt to argue that compatibilism is false. They claim that such a manipulated agent is not morally responsible for A-ing even if she satisfies the compatibilist conditions on moral responsibility when A-ing. They thus conclude that those conditions are insufficient for moral responsibility.

There have been a variety of twists and turns in the debate about manipulation cases and their associated arguments. An interesting development in the recent literature is Todd's (2012) Moral Standing Manipulation Argument (henceforth 'MSMA'). Normally, after presenting a manipulation case, incompatibilists ask whether a manipulated agent is morally responsible for her actions; they aim to elicit the response that such a person is *not* morally responsible. One strategy that compatibilists have to undermine

✉ Benjamin Matheson
Benjamin.matheson@philosophy.su.se

¹ Department of Philosophy, Stockholm Centre for the Ethics of War and Peace, Stockholm University, Stockholm, Sweden

this apparently intuitive response is to appeal to the reverse generalisation argument defended by Michael McKenna (2008, 2014) and John Martin Fischer (2011).¹ While incompatibilists claim that there are *no responsibility-relevant differences* between a manipulated agent and a merely causally determined agent and thus aim to generalise the non-responsibility intuition about the manipulated agent to merely determined agent, McKenna and Fischer argue that compatibilists can just generalise in the other direction. Because compatibilists think that a merely determined agent who satisfies the compatibilist conditions on moral responsibility *is* morally responsible, it follows that a manipulated agent (who also satisfies those conditions) is also morally responsible.² This isn't to be taken as an argument *for* compatibilism, but rather as a way to show that the non-responsibility intuition about the manipulated agent isn't sufficiently reliable to ground an argument against compatibilism. Hence manipulation arguments fail.

Todd's novel move seems to get around this style of response. Instead of focusing on the question of whether a manipulated agent is morally responsible, he focuses on the question of whether a *manipulator* can hold the person that she has manipulated responsible for her action. More precisely, he asks: can a manipulator *blame* a person she manipulates to perform some morally wrong action? Todd claims that it is counterintuitive to say 'yes' to this question. As he sees it, it is not permissible for a manipulator to blame the person she has manipulated.

I agree with Todd. It is counterintuitive for it to be permissible for a manipulator to blame a person for performing an action she (the manipulator) has manipulated her (the victim) to perform. Todd believes that the best way for the compatibilist to accommodate this intuition is to claim that the manipulator lacks the *moral standing* to blame her victim and most of his paper is spent attempting to rebut potential compatibilist attempts to defend this kind of reply. He concludes that this cannot be done satisfactorily – that is, compatibilists cannot find adequate support for the claim that a manipulator lacks the moral standing to blame her victim – and he takes this to support his contention that the *best explanation* for why the manipulator cannot blame her victim is that incompatibilism is true. In other words, the manipulator cannot blame her victim because the victim is not morally responsible due to being causally determined, and not because the manipulator lacks the moral standing to blame. Hence compatibilism is false.

In this paper, I argue that the MSMA does not refute compatibilism. In §1 I outline the MSMA in more detail. I then offer three ways to undermine the MSMA. In §2 I argue that the manipulators in the scenarios that support the MSMA satisfy Cohen's bad faith condition. I show that the MSMA only seems resist this reply because Todd uses *God* – an omnipotent, omniscient, and morally perfect being – as his manipulator. I argue that many compatibilists can reject this version of the argument because it appeals to a scenario that they do not think is conceptually possible – viz., a scenario in

¹ The reverse generalisation argument is a version of what McKenna (2008: 243) calls a 'hard-line reply'. It is a hard-line reply because it argues that the intuition that a manipulated agent is not morally responsible is (at least) unreliable, and so cannot support an argument against compatibilism. Other styles of hard-line reply have been defended by Frankfurt (2002: 28), Arpaly (2006), Talbert (2009), Russell (2010), Khoury (2013), and Matheson (2014)

² McKenna (2008, 2014) worries that it begs the question to start with the compatibilist's intuition about merely determined agents, so he generalises a neutral stance according to which *it is not clear that the merely determined agent is not morally responsible*. This, he thinks, is enough to undercut the incompatibilist's argument.

which God causally determines an agent to perform a morally wrong action. And I argue that non-morally perfect manipulators all satisfy the bad faith condition because of the mode of manipulation at issue. In §3 I argue that the manipulator in the scenario that supports the MSMA may satisfy Cohen's (2006) complicity condition on lacking the moral standing to blame. In §4 I argue that there is another condition that undermines moral standing to blame beyond those that Todd considers – namely the manipulation condition. In short, I argue that manipulators cannot blame their victims for *A-ing* *because* they have manipulated them to *A*. I show that this condition is not ad hoc by appealing to cases of manipulated non-determined agents and cases of everyday interpersonal manipulation.

1 The Moral Standing Manipulation Argument

Let's first consider the manipulation case that Todd uses to support his argument

Diana creates 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 [call him Ernie] 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 judgement, thereby bringing about event *E*. (Mele 2006: 188, as quoted in Todd 2012: 3)³

Todd holds that Ernie is not morally responsible for *A-ing* or for any of his actions because he has been manipulated by Diana – a minor goddess – to perform those actions. The catch is, of course, that it may be plausibly stipulated that Ernie satisfies all the leading compatibilist conditions on moral responsibility.⁴ Thus, as Todd sees it, those conditions are all insufficient for moral responsibility.

As Todd (2012: 4) notes, many compatibilists are inclined to resist this sort of manipulation case by simply denying that Ernie is not morally responsible. I include myself among such compatibilists. As far as I can see, there is nothing responsibility-undermining about the mode of manipulation that Ernie is subject to. (Although I don't *just* deny that Ernie is not morally responsible; I think there are arguments that show that he is, such as those defended by McKenna (2008, 2014) and Fischer (2011), but

³ Mele (2008: 278–280) claims that his scenario is not a manipulation case, but rather an 'original design' case. Likewise, Fischer (2011: 271) calls it an 'initial design' case. Todd (2012: 2, n.1), however, assumes that original/initial design is a form of manipulation, so I will follow Todd in holding that the Diana/Ernie case is a manipulation case.

⁴ It won't matter for my purposes what these compatibilists conditions are because nothing in my argument hangs on their specifics. Roughly, compatibilists have an option between a 'historical' account (e.g. Mele 2006; Fischer and Ravizza 1998; Haji 2013) or a 'non-historical' account (e.g. Frankfurt 1971, McKenna 2004, Arpaly 2006, Talbert 2009, Russell 2010, Khoury 2013, Matheson 2014). According to the latter sort of account, an agent's moral responsibility depends only on non-historical facts about the agent at the time of action. According to the former sort of account, an agent's history (in some sense) is also relevant to her moral responsibility.

let's set aside these arguments here.) But as Todd (2012: 4) notes, his argument is an effort to press compatibilists in a different way and therefore to open up a new 'dialectical pathway' with which he seeks to undermine compatibilism.

Todd starts by modifying the Diana/Ernie scenario. While Mele doesn't specify what sort of action Ernie is manipulated to perform, Todd asks us to suppose that Ernie has been manipulated by Diana to murder Jones. As noted above, Todd doesn't focus on the question of whether Ernie is morally responsible, but rather on a different question – namely, 'can *Diana* blame Ernie? That is, having once determined Ernie to kill Jones, can Diana later blame him for doing so?' (2012: 4). Todd claims that it is counterintuitive that we may blame Ernie, but that 'it seems much more counterintuitive to suppose that Diana may blame Ernie for what she has determined him to do' (2012: 4). As Todd implicitly concedes, there is no progress to be made on the question of whether it is counterintuitive that Ernie *is* responsible, so we can set that aside here and simply focus on whether it is counterintuitive for it to be permissible for Diana to blame Ernie.

As I've said, I agree that it is counterintuitive for it to be permissible for Diana to blame Ernie for killing Jones given that she manipulated him to do so, and I agree it would be damaging for compatibilism to imply that it is permissible for Diana to blame Ernie. But, as I'll argue, this is not an implication of compatibilism – that is, compatibilists can explain why Diana is not permitted to blame Ernie. This is perhaps why at this point in his argument Todd modifies the Diana/Ernie scenario in a second way. Instead of Diana manipulating Ernie, Todd asks us to suppose that *God* is Ernie's manipulator. Todd justifies this move by claiming that 'unless we make a wide range of stipulations, it may turn out that Diana cannot blame Ernie for reasons not relevant to the issue at hand' (2012: 4). For example, it might turn out that Diana cannot blame Ernie because she 'is a rather wicked [and] conniving goddess' (2012: 4). Thus Todd, in effect, upgrades from a quite powerful and knowledgeable being to an all-powerful, all-knowing, and perfectly good being. As we'll see in §2, one of reasons why Diana is not permitted to blame Ernie is in fact relevant to the issue at hand – but let's set this point aside until then.

Todd claims that, just as it is counterintuitive for Diana to blame Ernie, it is also counterintuitive for *God* to blame Ernie for murdering Jones. He compares God to an author and the agents God determines to characters in a story that God has written. He then asks, "How could it be appropriate for God to blame [determined agents] for (so to speak) *perfectly performing* the roles he assigned to [them]?" (Todd 2012: 5). I agree with Todd that it is counterintuitive for it to be appropriate (that is to say, permissible) for any manipulator to blame agents for performing actions they (the manipulator) have manipulated them (the agents) to perform. The intuitive plausibility of the claim that a manipulator cannot blame her victim(s) is used to support the first premise of Todd's argument:

1. On theological determinism [the thesis that God creates a causally deterministic world], God cannot blame us for the wrong actions we perform, even if we meet all the compatibilist conditions on being morally responsible with respect to performing them.
2. The best explanation for the truth of (1) is incompatibilism.

So,

3. Incompatibilism Is True. (Todd 2012: 5)

Todd notes that he thinks compatibilists will agree with him that (1) is true. I am one such compatibilist. However, I think that (2) is false. I will argue that there are better explanations of (1) and modified versions of (1). In §2 I will argue that because (1) is supported by a scenario that is not conceptually possible (i.e. one where God causally determines us to do morally wrong things), the best explanation for (1) is actually that God cannot causally determine us to perform morally wrong actions. Of course, (1) can be modified to make no reference to God. But then I argue that incompatibilism is still not the best explanation of a modified (1). Rather, the best explanation for why a manipulator is not permitted to blame her victims is that the manipulator lacks the moral standing to do so. I argue that non-morally-perfect manipulators in the scenario that supports the MSMA satisfy Cohen's bad faith condition on lacking the moral standing to blame. In §3 I argue manipulators in the scenario that supports the MSMA also satisfy Cohen's complicity condition. I consider and reject Todd's argument against such a reply. In §4 I argue that manipulators satisfy another condition on lacking the moral standing to blame – namely, the *manipulation condition* – which is expressed by the complaint: you manipulated me! Unlike my response in §2, my responses in §3 and §4 also apply to God.

My response to the MSMA can be distinguished from another style of response. Mine agrees with the underlying intuition that supports the argument, and therefore accepts the challenge the argument presents. Such a style of response thus attempts to show how compatibilists can accept this intuition without having to reject compatibilism. Another other style of response rejects the underlying intuition and thereby denies the challenge. Such compatibilists simply deny (1) on the grounds that God is permitted to blame us (including Ernie) for the wrong actions we perform when we satisfy the relevant conditions on blameworthiness *because* we are blameworthy for so acting. Such a compatibilist believes that it is irrelevant what causal role God had in the production of our actions. Since we satisfy the relevant conditions, God can permissibly blame us. Such a compatibilist response might disregard considerations about moral standing to blame – that is, such a compatibilist might hold that moral standing never affects appropriate blaming. According to this kind of view, someone who is blameworthy can be blamed *by anyone* simply in virtue of the fact that they are blameworthy.

This is certainly a hard-line to take, and one that is too hard for me to take. I find it highly counterintuitive that a manipulator, including God, could permissibly blame Ernie (or anyone else) for a morally wrong action they manipulated Ernie (or anyone else) to perform. As I've said, I share the intuition underlying the MSMA, and so I see the force of Todd's challenge.

My overall strategy for responding to the MSMA is to propose three ways to undercut the argument while accepting that the argument presents a challenge to compatibilism. I now present the first of my three responses.

2 Bad Faith and the Importance of the Manipulator

2.1 Cohen on Bad Faith

Following Cohen (2006), Todd considers the following two conditions that undermine the moral standing to blame: bad faith and complicity. Cohen's bad faith condition relates to a kind of hypocrisy. It seems plausible that those who are prolific thieves

often cannot legitimately blame others for stealing because they are the kind of people who steal. As Todd (2012: 7) writes, ‘what is objectionable about (at least some cases of) hypocritical blame is that the hypocrite represents himself as caring about the moral values at stake, but his behaviour in doing the very same things suggests otherwise. The worry then is that the hypocrite is blaming in *bad faith*.’ The idea, I take it, is that blaming involves taking a kind of moral high ground and prolific thieves usually cannot do that when it comes to others who steal because they are not on the high ground with respect to that type of action. Thus, while an agent might be blameworthy for something (such as stealing), there might be other agents who cannot appropriately blame her for doing so.

To appropriately blame Ernie therefore seems to mean that you at least believe that Ernie’s killing of Jones is morally wrong. But a compatibilist might argue that because God causally determines the world to include Ernie’s killing of Jones, God cannot believe this action is wrong. God, after all, is morally perfect. So if God blames Ernie, he does so in bad faith – that is, he doesn’t really believe it is wrong. Todd responds to this line of argument in the following way. He says that God might want a *greater good* to arise from Ernie killing Jones. And God either intends for Ernie to kill Jones or he doesn’t. If God intends for Ernie to kill Jones, this apparently doesn’t mean he *approves* of Ernie killing Jones. If God doesn’t intend for Ernie to kill Jones, then Ernie killing Jones is merely a foreseen but unintended consequence of God trying to achieve the greater good he wants to achieve. Either way, God’s intentions are still apparently good (Todd 2012: 8). The bad faith objection thus seems to get no purchase. So Todd concludes that this attempted compatibilist response fails.

I will now argue that Diana satisfies the bad faith condition. I will then point out that using God as the manipulator ‘rigs the game’ because God (by hypothesis) is morally perfect and so cannot lack the moral standing to blame a blameworthy agent on pain of inconsistency. I then argue that compatibilists who reject classical monotheism need not accept that the scenario in which God manipulates Ernie to perform a morally wrong action is a conceptually possible one. Hence the argument fails with God as the manipulator.

2.2 Diana as Manipulator

Consider Todd’s argument with the original zygote scenario that features Diana instead of God. Compatibilists have a straightforward response to this variant of the argument: Diana lacks the moral standing to blame Ernie even though he is blameworthy for killing Jones, because she satisfies the bad faith condition on lacking the moral standing to blame him. Let’s unpack this.

Diana intentionally and knowingly manipulates Ernie to murder Jones. Someone who intentionally and knowingly brings about a morally bad outcome is likely to be someone who endorses morally bad things. So we have good reason to think that Diana satisfies the bad faith condition. Of course, as Todd points out, there is a difference between the person who murders someone and the person who manipulates that person to murder someone. He writes:

Bringing about conditions which causally determine someone to pick up a gun and shoot someone for selfish reasons is not the same thing as picking up a gun and shooting someone for selfish reasons. (Todd 2012: 6–7)

Todd makes this point while arguing that God does not satisfy the bad faith condition. Todd's purpose is to show that God does not do the same thing as Ernie, and so it's not obvious that God satisfies the bad faith condition. It is true that neither God nor Diana do the same thing as Ernie. However, the bad faith condition does not concern what a person has actually done but, as Todd later discusses, rather concerns what people approve of or endorse. We can test this by asking: could a person who approves of murder permissibly blame another person for murdering someone? It seems not. To blame someone for *A-ing* is to represent yourself as believing that *A-ing* is bad, and the former person clearly does not think that murder is bad because she approves of murder. Compare them to a person who has killed in self-defence. Despite having done something similar to a murder, the self-defender can still permissibly blame a murderer for what she has done because the self-defender need not approve of murder.

Now, as I'll discuss in §2.4, it's clear that God cannot approve of murder because God is morally perfect. But this is not the case with Diana. My claim is that Diana in virtue of manipulating Ernie *in the manner* she does cannot help but approve of murder when she manipulates him. As I argue in §4, though, it remains open for more everyday manipulators to manipulate someone to do something morally wrong without approving of that moral wrongdoing. But Diana is no everyday manipulator. She uses her power and knowledge of all facts about the past and the laws of nature to causally determine everything Ernie ever intends and does. Given the *kind* of manipulation that Diana uses to get Ernie to kill Jones, she cannot avoid approving of his action.

This is partly because Diana *simulates* having Ernie's bad intention that leads to him murdering Jones. While this is not explicitly part of the case as Mele or Todd present it, it seems to be implied by the details of the case (cf. Barnes 2015). Diana does not go through Ernie's actual thought process (since only he can go through that). But by meticulously setting things up the way she does, Diana effectively goes through that thought process before he even has and so she has an understanding of what that intention is like. As Todd notes, the manner in which Diana manipulates Ernie makes Diana relevantly like an author. Authors also simulate what it is like to have bad intentions. It is something that they must do in order to write compelling stories. Most authors, however, do not go into the detail that Diana does. She writes every detail of Ernie's life, every thought, every intention, every action, and so on. Even if we were to imagine such an author, the author would seem to *only* simulate having bad intentions, and need not approve of the content of those intentions. For example, an author who writes about murder need not approve of murder. This is true, but there is a difference between Diana and such an author: while such an author writes stories with agents who have bad intentions and who do bad things, Diana does that *and* she knowingly actualises those stories. So while simulating bad intentions does not itself result in one approving of the content of those intentions, my claim is that simulating bad intentions and knowingly making those intentions actual in the manner that Diana does in fact results in her approving of those intentions.⁵

⁵ I accept that reasonable people may disagree about this claim. I have not offered a knockdown argument for why Diana must simulate Ernie's malicious intention. Those who disagree with this claim will therefore be unable to accept the first of my three replies to the MSMA. But because my other two replies do not rely on the claim, those who disagree may still be able to endorse one of my other replies. Thanks to an anonymous referee for pressing me on this point.

Because Diana actualises her story with the full knowledge of what that involves, she cannot help but approve of Ernie murdering Jones when she manipulates Ernie to do so. She might also be blameworthy for Jones's murder. That depends on whether she satisfies the relevant conditions. But we need not determine whether she is blameworthy or not, because approving of Jones's murder when she manipulates Ernie is sufficient for her to satisfy the bad faith condition with respect to blaming Ernie. So Diana lacks the moral standing to blame Ernie. Compatibilists therefore have an adequate explanation for why Diana is not permitted to blame Ernie. And note that while the original case features Diana intending for Ernie to kill Jones, the point goes through even if we stipulate that Diana doesn't intend for this to happen. The strength of this point lies with the fact that Diana knowingly manipulates Ernie to murder Jones. So the MSMA supported by the original Diana/Ernie case does not show that compatibilism is false.

2.3 For the Greater Good

But what if Diana (let's call her Diana* to avoid confusion) manipulates Ernie to achieve some greater good? Perhaps Diana* was presented with a range of possible options for achieving this greater good, and all but one of them would have resulted in lots of people dying. The option that will result in the fewest people dying is one in which she manipulates Ernie in the manner at issue to kill Jones. So Diana* opts to do this. But Ernie must also believe his killing of Jones is morally wrong.⁶ So let's suppose that from Ernie's perspective killing Jones is a bad thing to do and that Ernie kills Jones as a result of a set of bad intentions. It therefore seems that Ernie, according to at least some compatibilist views, is blameworthy for killing Jones.⁷ But it seems counterintuitive for it to be permissible for Diana* to blame Ernie for killing Jones. After all, Diana* manipulates Ernie to do that. And it seems that the explanation offered in §2.2 will not apply to Diana*. She, after all, manipulates Ernie for the greater good, so it seems that she need not approve of Jones's murder when she manipulates Ernie. This scenario therefore seems to support the MSMA. But, as I will now argue, Diana* must approve of Ernie killing Jones to some extent when she manipulates Ernie to do so, and so she still satisfies Cohen's bad faith condition.

Remember that she, as a result of how she manipulates Ernie, has effectively gone through Ernie's thought processes leading up to the murder before he even has. Diana* simulates what it is like to have and to act from bad intentions. Because she actualises those bad intentions with her knowledge of the past and the laws of nature, this results in her endorsing the content of those intentions. The fact she was willing (at least temporarily) to endorse bad intentions and the morally wrong actions that resulted from them for a greater good might make her praiseworthy,

⁶ Several authors – such as Capes (2012), Haji (1998), and Khoury (2011) – have argued blameworthiness requires that the agent believes she is doing wrong.

⁷ Note that if one can only be blameworthy for actions that are objectively morally wrong, then Ernie might not come out as blameworthy because it's not clear that, on the picture that Todd presents, that Ernie murdering Jones is objectively morally wrong. If God or Diana* is justified in getting Ernie to murder Jones, then it might be that Ernie's action is not morally wrong, even if Ernie believes it is morally wrong. This raises many complex issues that cannot be dealt with adequately here, so I will set this worry aside.

but this doesn't mean that she has the standing to blame Ernie.⁸ It remains that she endorses or approves of Jones's murder when she manipulates Ernie. She thereby satisfies the bad faith condition on lacking the moral standing to blame because of how she brought about the greater good. Compatibilists thus have an adequate explanation for why Diana* cannot blame Ernie. Hence (2) is false, and so this variant of the MSMA is also unsuccessful.

2.4 God as Manipulator

We can now see why Todd was forced to switch Diana out for another manipulator: the argument does not work with Diana (or Diana*) as the manipulator. Instead of Diana, Todd uses God – an omnipotent, omniscient, and morally perfect being. While it makes sense to say that Diana lacks the moral standing to blame a blameworthy agent (because she satisfies the bad faith condition), it does not make sense to say that God lacks the moral standing to blame a blameworthy agent. Why? Well, because God is a *morally perfect* being and as such cannot – on pain of contradiction – lack the moral standing to blame a blameworthy agent. If God did lack the moral standing to blame a blameworthy agent, then God would not be morally perfect. Given that God cannot lack the moral standing to blame Ernie even though Ernie satisfies the compatibilist conditions on blameworthiness, it might then seem that the best explanation of why God can't blame Ernie is that Ernie is not, in fact, blameworthy for killing Jones. In other words, incompatibilism is true, and hence compatibilism is false.

Todd's choice of manipulator is problematic, however. While it might be uncontroversial that God (as defined) is conceptually possible (though I think this is questionable), it is not uncontroversial that God could manipulate someone to perform a morally wrong action. That might be a comfortable assumption for classical monotheists, but not all compatibilists are classical monotheists. Some compatibilists are classical monotheists, of course, so they will likely agree with this assumption. I set such compatibilists aside in the rest of this section; my first response will offer them no assistance in response to the MSMA. Rather, I shall focus on those compatibilists who reject classical monotheism. Many such compatibilists will reject Todd's assumption because they believe that God's existence is incompatible with the existence of evil.⁹ The so-called logical argument from evil does not give such compatibilists reason to believe that God is conceptually impossible. But it does entail that God does not exist in possible worlds that contain evil, and the world under consideration contains evil (namely Jones's murder). Even more minimally, the compatibilist need only maintain that it is highly controversial whether or not God can exist in worlds that contain evil. This is enough to generate a compatibilist response to the original MSMA.

I agree, of course, that compatibilists must accommodate scenarios that involve *uncontroversially* conceptually possible beings – such as nefarious neuroscientists,

⁸ By endorsing a morally wrong action, there is a sense in which Diana* gets 'dirty hands'. Some (e.g. Walzer 1973) have argued that sometimes what we think is morally indefensible is sometimes permissible given the extremity of certain situations. Even though we must do what is morally wrong, it remains morally wrong.

⁹ Todd (2012: 9) notes that his argument touches upon the problem of evil. This part of his argument raises many questions. Unfortunately, an adequate discussion of all these issues is beyond the scope of this paper.

advanced aliens, or even minor deities – but they do not need to accommodate scenarios that are not uncontroversially conceptually possible. And the scenario involving a perfect being creating something imperfect – such as God manipulating Ernie to kill Jones – is not uncontroversially conceptually possible, as the controversy surrounding the logical argument from evil provides us evidence of. For the incompatibilist to make a compelling argument against compatibilism they need to use uncontroversial manipulators, such as imperfect deities, nefarious neuroscientists, and advanced aliens. While these beings are all fanciful and farfetched, there is a widespread agreement that such beings are conceptually possible – we can all see how they might exist and how they might manipulate someone – so they can be used in scenarios to ascertain the conceptually necessary and sufficient conditions on moral responsibility.

So compatibilists can accept that (1) is true, but argue that (2) is not the best explanation of (1). Rather, the best explanation of (1) – that is, for why God cannot blame us for morally wrong actions he causally determines us to perform – is that God cannot causally determine us to perform morally wrong actions (just as he cannot do anything else that is conceptually impossible). Hence this variant of the MSMA is unsuccessful.

Compatibilists are not here denying that God exists. Rather, they are simply claiming that God cannot do certain things. In particular, he cannot manipulate people (whether through causal determination or otherwise) to perform morally wrong actions. The same considerations apply to merely morally perfect beings. Just as compatibilists can deny that the scenario in which God manipulates Ernie is conceptually possible, compatibilists can also deny that the scenario in which a morally perfect being manipulates Ernie is conceptually possible. This does not mean that compatibilists must accept that a morally perfect being is conceptually impossible. It only means she holds that a morally perfect being cannot do certain things, which is uncontroversial. As I argued in §2.2, because of how Diana manipulates Ernie she must also minimally approve of the content of his intentions when she manipulates him. I don't see how a morally perfect being could approve of murder; any being that approves of murder, even temporarily, is not a morally perfect being in my book. So, while a morally perfect being might be conceptually possible, it is not conceptually possible for a morally perfect being to manipulate Ernie in the manner required to murder Jones to get the MSMA off the ground. This is all the compatibilist must show to undercut this version of the MSMA. So the MSMA is unsuccessful whichever kind of manipulator is used.

3 Complicity

In this section, I argue that manipulators (in the scenarios that potentially cause problems for compatibilism) may satisfy the complicity condition. Unlike my response in §2, this response is open to all compatibilists. The complicity condition involves claims like “you're involved too”. For instance, equal co-conspirators plausibly cannot blame each other for doing the thing they conspired to do because they are all equally involved in the production of the outcome. The attempted compatibilist explanation of (1) that appeals to this condition, then, says that God

is not permitted to blame Ernie because God is involved in the killing of Jones. In other words, the compatibilist contends that God is complicit in Jones's death, and so God cannot blame Ernie on that basis.

Todd accepts that God might be *morally responsible* for Ernie killing Jones, but argues that being morally responsible for something is insufficient for being complicit in that something coming about. Given that God is not complicit, he doesn't lack the moral standing to blame Ernie for that reason. And if we can't find a reason why God lacks the moral standing to blame Ernie, we must explain why God cannot blame Ernie in some other way – and Todd's other way is to say that Ernie is not blameworthy because incompatibilism is true. Todd supports his point by appealing to the following case.

Jonas is a Nazi commander secretly opposed to the Nazi regime. Jonas uses his position to save lives where possible, but sometimes must do things to maintain his cover. One day he notices a break out from the prison where he works and he orders Thomas, a subordinate of his, to check it out, as protocol dictates. Jonas hopes Thomas will do the right thing and let any prisoners go, but unfortunately Thomas doesn't. Todd claims that while Jonas is morally responsible for the prisoners not escaping (because he ordered Jonas to investigate) Jonas retains the standing to blame Thomas for not letting the prisoners escape. According to Todd, Jonas has the standing to blame Thomas because he (Jonas) is morally opposed to the Nazi regime and because it is apparently not Jonas's fault (in the blame-implying sense) that the prisoners did not escape. That is, Jonas retains the standing to blame Thomas because, according to Todd (2012: 11), Jonas does not endorse or approve of the Nazi regime or its values and he is apparently not blameworthy for what Thomas does. If we agree that Jonas is not blameworthy, then this shows that one can be morally responsible for an outcome without losing one's moral standing to blame someone else for that outcome. Because (according to Todd) God is only morally responsible for what Ernie did, God's being morally responsible for Ernie killing Jones cannot be the reason why God seems to lack the moral standing to blame Ernie. Given that it seems that complicity does undermine moral standing, merely being morally responsible for X is insufficient for being complicit in bringing about X. Consequently, God's seeming lack of moral standing to blame Ernie apparently cannot be because God is complicit in what Jones did. And, again, if compatibilists cannot find a reason that God lacks moral standing to blame Ernie, the fact it is impermissible for God to blame Ernie may then be that Ernie is not blameworthy for killing Jones because compatibilism is false.

Todd's attempted rebuttal of this response relies on us agreeing with his intuitive judgement about the Jonas/Thomas case. We must agree that Jonas doesn't lose the standing to blame Thomas for complicity reasons, despite the fact that Jonas ordered Thomas to do what he did. I'm not sure what my intuitions are about this case. It's not clear to me that Jonas does have the moral standing nor is it clear that he lacks the moral standing. It is only because it apparently seems intuitive that Jonas retains the moral standing to blame Thomas that we have reason to think that being morally responsible for X is insufficient for complicity in bringing about X. Because it is not clear to me what my intuitions are about this case, I do not see why compatibilists cannot appeal to this line of defence. But this is of course a different story for those who agree with Todd's assessment of this case. Should such people trust their

intuition, though? If they can't trust this intuition, then they have no reason to think that being morally responsible for X is insufficient for being complicit in bringing about X, and so they have no reason to think that God's being morally responsible for bringing about Ernie's killing of Jones doesn't make God complicit in Ernie's killing of Jones. In what follows, I will present a reason that suggests that if you share Todd's intuition, you shouldn't trust that intuition.

To see this reason, we must first appreciate that there are standing conditions on all negative attitudes we might have towards others and that some of these conditions seem independent from one another. Take, for example, contempt. It seems clear that a committed Nazi deserves our contempt. Thomas is a committed Nazi. So, he deserves contempt. I will now argue that just because one has the standing to have contempt for a person it does not follow that one also has the standing to blame that person. I will then argue that because it seems intuitive that Jonas has the standing to have contempt for Thomas that we may have mistakenly inferred that he also has the standing to blame Thomas.

Contempt is a *person-focused* attitude (Mason 2003) and so doesn't obviously require that a person is blameworthy for having their contemptible traits. All that matters for deserving contempt is that a person is contemptible, and a person being contemptible depends on her having contemptible traits. That someone deserves blame for their traits is a further question, which arguably depends on whether they brought about or developed those traits. Because Thomas is not only blameworthy for not letting the prisoners escape but is also a committed Nazi, he deserves blame as well as contempt: blame for his action and contempt for being a committed Nazi. The question is whether Jonas has the standing to blame Thomas. The main thing to note is that blame has a different *object* to contempt. While contempt is person-focused, blame is *action-focused*. We have contempt for a person (or hold them in contempt) for who they are (i.e. what traits they have), but we blame a person for what they have done. Given their different objects, it doesn't automatically follow that because one has or lacks the standing to blame that one has or lacks the standing to have contempt – and vice versa. It seems plausible that Jonas has the standing to have contempt for Thomas. Why? Well, because contempt takes traits as objects, the standing conditions for having contempt for a person plausibly depend on one's traits. A committed Nazi lacks the standing to have contempt for another Nazi for being a Nazi because the committed Nazi is a committed Nazi. But, by hypothesis, Jonas has non-Nazi traits and Thomas has Nazi traits (to put it roughly). It thereby seems plausible that Jonas has the standing to have contempt for Thomas. This is true regardless of what Jonas orders Thomas to do.

Notice that the following general claim is true when we consider this case: Jonas has the standing to feel and express certain negative attitudes about Thomas. However, because we see that Jonas has such standing, I submit that Todd and those who share his intuition may have inferred Jonas also has the standing to blame Thomas. But this doesn't follow. Complicity seems to undermine standing to blame because complicity and blame are *causal* phenomena: they are about things we have caused (though they of course require more agential involvement than mere causal responsibility). But complicity doesn't seem to affect one's standing to have contempt for another, because contempt is about a person (i.e. her traits) rather than what she does. This gives us reason to be suspicious about the intuition that Jonas has the standing to blame Thomas. It might seem intuitive that Jonas has the standing to blame Thomas because Jonas has

the standing to have contempt for Jonas. But, as argued, this doesn't automatically follow.

If Jonas does not have the standing to blame Thomas for complicity reasons, then it seems that God (or Diana) might also not have the standing to blame Ernie for complicity reasons. There is a lot more discussion about standing and negative attitudes that I lack the space to delve into here. For instance, it might well be that while God (or Diana) has the standing to blame Ernie, he (or she) lacks the standing to have contempt for him because he (or she) created Ernie exactly as he is. Given this, my point is far from decisive. Despite this, compatibilists have a *prima facie* principled response to the MSMA: God (or Diana) lacks the standing to blame Ernie for complicity reasons. The underlying point is that our intuitions about standing may be muddled and unreliable because there are many more factors than we realised in the mix. It is just not clear we can neatly separate our intuitions about standing for the various negative attitudes that people like Jonas or Ernie are a deserving target of. My point at least shifts the burden onto Todd to develop his argument against this alternative explanation of why God (or Diana) lacks the moral standing to blame Ernie. Until my point is countered, it seems fair for compatibilists to hold that God (or Diana) cannot blame Ernie because God (or Diana) lacks the moral standing to blame Ernie for complicity reasons.

4 You Manipulated me!

I now present a third reply to the MSMA, one which is also open to all compatibilists. Todd considers various ways compatibilist might support the claim that God lacks the moral standing to blame by appealing to and modifying Cohen's conditions on when moral standing to blame is undermined. But we have no reason to think that his conditions are exhaustive of the conditions that undermine the moral standing to blame. I hereby propose a new condition on when moral standing to blame is undermined – namely *the manipulation condition*. It is expressed by the complaint “you manipulated me!”

Throughout this paper (and throughout Todd's), we've seen that this explanation has been invoked to explain why it is inappropriate or impermissible for a manipulator to blame her victim. We ask: why is it inappropriate for a manipulator to blame her victim? Answer: because the manipulator *manipulated* her victim. Todd's presumes that this explanation is not foundational – i.e. when we think manipulation undermines moral standing to blame, it must be because another factor is playing this role. But why? There's no argument here. As I've said, it's presumed. I will now argue that “you manipulated me!” is a non-ad hoc and distinct category on lacking the moral standing to blame – that is, we have independent reasons for taking it to exist and it need not reduce to another condition on lacking the moral standing to blame.

First, note that Todd sometimes moves smoothly between talk of ‘manipulation’ and talk of ‘causal determination’. However, being causally determined is neither necessary nor sufficient for being manipulated.

It is not necessary because it is possible to manipulate an agent without causally determining her. We might imagine cases where Diana manipulates Ernie* in a non-deterministic way. Perhaps Diana allows that there is a tiny chance that Ernie* will not kill Jones. But it still seems that Diana cannot blame Ernie* for actually killing Jones.

Indeed, although this is more controversial and my point goes through without it, I think that it is plausible that Diana is not permitted to blame Ernie* even if he satisfies all the libertarian conditions on being morally responsible. Even if libertarians think that Ernie* *is* morally responsible, they do not need to accept that Diana, because she manipulated him to do so, is permitted to blame Ernie* for killing Jones. It strikes me that it is just as counterintuitive for Diana to blame Ernie* as it is for Diana to blame Ernie.

It is not sufficient because merely causally determining an agent does not mean you have manipulated her. If Diana created Ernie's zygote with no knowledge of what he would do, then she would still causally determine his whole life. However, it doesn't seem that she has manipulated Ernie in any way. Indeed, it seems fine for this Diana to blame Ernie. After all, she didn't manipulate him (and she plausibly doesn't satisfy any other conditions on lacking the moral standing to blame him either).

I propose, then, that the MSMA is unsuccessful in all its variations simply because manipulators cannot blame their victims for the wrongdoing in question purely in virtue of the fact that they manipulated their victim to perform that wrongdoing. It is because the manipulator stands in a particular relation to her victim that explains why she cannot blame her victim, and that relation is *manipulator to manipulatee* – what we might call the 'manipulation relation'. This is not ad hoc – that is, the motivation for invoking this new condition is not simply to get the compatibilist the result she requires. If it only applied to compatibilists, then *you manipulated me!* would be tantamount to saying *you causally determined me!* But since it applies to non-determined agents, it's not just about causal determination. And causally determining someone, as I argued above, does not undermine your moral standing to blame that person. So, causal determination is neither necessary nor sufficient for manipulation. Hence, manipulation is not an ad hoc condition on lacking the moral standing to blame.

We can support this further by appealing to cases of interpersonal manipulation – that is, the everyday cases that do not involve outlandish circumstances and hypothetical beings. Suppose Lorraine wants Randy to hit Ken. She might start talking to Randy about Ken in more and more negative ways. She might even bring up things Ken has said about Randy out of context, or flat out lie about what Ken has said about Randy. Slowly over time Randy's perception of Ken changes; Randy comes to dislike Ken and eventually hits him. Lorraine has succeeded. Has Lorraine undermined Randy's free will in during this process? It's not clear how. Indeed, it seems that whatever you think free will amounts to that Randy retains it during this process. And he plausibly satisfies other conditions on being morally responsible, such as epistemic conditions (e.g. he knows the moral status of his actions), throughout the manipulation process.¹⁰ So we have no reason to think Randy is not morally responsible for hitting Ken. But is it permissible for Lorraine to blame Randy for hitting Ken? It seems highly counterintuitive to hold that this is permissible. After all, she manipulated Randy to hit Ken!

One might worry that my proposed condition is not foundational – that is, it is actually one of Cohen's two conditions. Consider first his complicity condition – i.e.

¹⁰ It's worth noting that many have typically failed to acknowledge that their theories of free will and moral responsibility must accommodate people acting freely and being responsible when they are manipulated in this way. Indeed, I think that exposing this fact will ultimately help to undermine manipulation arguments – but that's a discussion for another occasion.

“you’re involved too”. The worry is that manipulating someone amounts to the manipulator being involved the outcome. I argued in §3 that God or Diana may satisfy this condition, but I think they may satisfy this condition on the basis of the *way* they manipulated Ernie rather than simply because they manipulated him. I shall now argue that Lorraine does not satisfy this condition, thereby vindicating my claim that the manipulation condition does not boil down the complicity condition.

Cohen (2006: 123) notes that the complicity condition involves points like “you made me do it” and “you started it”. Could Randy claim that Lorraine made him hit Ken or that she started the whole thing? Let’s start with the latter. Does Lorraine “start things” in this context? It’s not clear what this charge amounts to. She only ever talked to Randy while intending that he do things. But we do that all the time without making ourselves complicit. Suppose I ask you to pass the salt but I intend for you to throw salt someone else’s eyes, and then you do throw salt in someone else’s eyes. There’s a sense in which I started things by asking you to pass the salt, but it isn’t a sense in which I’m complicit in what you do (though we might think I satisfy the bad faith condition, given my intention – more on this shortly). The same is true with Lorraine. She starts things, but not in a way that makes her complicit. She never told Randy to hit Ken or gave Randy a plan or anything like that; maybe if she’d done that then she’d be complicit, but she hasn’t. Indeed, she’s been careful to avoid “starting” anything, even though she is actively manipulating Randy. This is part of what makes interpersonal manipulation so devious and so hard to detect.

It’s not the case that Lorraine made Randy do anything either. The whole point of her manipulating Randy in the way she did was so that he would act *of his own free will*. She could have gotten a gun and forced him to do it at gunpoint and thereby undermined his free will, but that would have clearly made her complicit – and indeed may have made her solely blameworthy for the outcome. Instead, she worked slowly and patiently to achieve her desired end. But she didn’t ever make Randy hit Ken. She never made Randy do anything. She was careful not to do that. She was careful not to be forceful. In short, she was careful not to become complicit. All she did was arrange circumstances and suggest things so that Randy would eventually do what she wanted. So it’s not true that she made Randy hit Ken. It was Randy who did that. But this doesn’t mean that Lorraine can legitimately blame Randy. Why? I claim it is because *she manipulated him*. And that’s enough to undermine her moral standing to blame Randy for hitting Ken.

Some might worry that Lorraine satisfies Cohen’s bad faith condition. It’s certainly true that a lot of real life manipulators will also satisfy this condition, because manipulation is typically a nefarious business conducted by those with malicious intentions and dubious moral characters. But it is not necessary that they do. It is possible Lorraine was actually trying to achieve some greater good by manipulating Randy to hit Ken. Perhaps her manipulating Randy was the only way to stop something much worse from happening. Perhaps Ken was going to hurt himself or others unless he could fight someone, and he is only able to fight Randy. Given that Lorraine does not have the power or knowledge that is even close to Diana’s, it is not the case that in manipulating Randy to hit Ken that Lorraine must simulate Randy’s intentions. Some manipulators might do this, but that doesn’t mean that they all must. It is possible for Lorraine to manipulate Randy while being entirely ignorant of his intentions. So there is nothing that comes close to forcing her to approve of or to endorse

Randy's bad intentions and the resulting action. Remember that Diana must approve of Ernie's intentions (and the resulting action) when she manipulates him because she simulates his intention and because she knowingly actualises his bad intention. This doesn't typically occur in cases of everyday interpersonal manipulation. I suspect it is these everyday cases that might make it seem possible for a being like Diana to manipulate Ernie whilst not approving of what Ernie does. This is possible for normal people manipulating others in humdrum ways; it is not possible for such god-like beings and the kind of manipulation that they are described as using to get the MSMA off the ground.

Of course, if it were possible for Diana and her ilk to manipulate people without coming to approve of their actions, then my point in this section would go through more easily. That is, it would be clear that manipulators need not endorse the intentions and actions of those they manipulate, and so it would be clear that they need not satisfy the bad faith condition. The incompatibilist therefore faces a dilemma: either all manipulators must approve of intentions and actions they manipulate others to have and to perform or they don't. If they do, then all manipulators would blame in bad faith. If they don't, then it remains that all manipulators cannot blame those they manipulated purely because they have manipulated them. Either way, manipulators cannot blame those they manipulate to perform morally wrong actions.

We therefore have another reason why God, Diana, or any other manipulator is not permitted to blame a person they manipulate to perform a morally wrong action – namely because they manipulate that person to perform that action.¹¹

5 Conclusion

In this paper, I have argued that the MSMA does not undermine compatibilism. I presented three ways to undermine this argument. The first way argues Todd's argument only seems *prima facie* successful because it uses God – a morally perfect manipulator. I then argued that given that it is controversial that it is conceptually possible for God, or any other morally perfect being, to manipulate Ernie, or anyone else, to perform a morally wrong action that compatibilists need not respond to arguments supported by such scenarios. That is, compatibilists can reject arguments supported by such scenarios on the basis that they find them to be conceptually impossible. And I argued that the argument does not work with less than morally perfect manipulators because all such manipulators satisfy the bad faith condition on lacking the moral standing to blame. The second way argues that compatibilists can reasonably contend that complicity undermines the standing to blame and that God and Diana are complicit in Ernie's killing of Jones, so they cannot appropriately blame Ernie for complicity reasons. The third way says that manipulators lack the standing to blame their victims simply because they have manipulated them.

¹¹ See Russell (2010) for a defence of compatibilism from manipulation arguments that starts from the observation that manipulators cannot blame their victims (though he doesn't use the terminology of moral standing).

These responses are independent and compatible. So compatibilists can endorse one, two, or all three of them. I've argued that a manipulator who manipulates someone in the manner that Diana or God manipulates Ernie satisfies three conditions on lacking the moral standing to blame Ernie: complicity, bad faith, and manipulation. Some compatibilists, especially those who are classical monotheists, will likely reject my first reply, but can endorse my second and third replies. These replies might have other implications for them – e.g. God cannot blame those he manipulates – but such compatibilists can still maintain that it is conceptually possible that God causally determine us to perform morally wrong actions, so such compatibilists can still be classical monotheists. Other compatibilists might reject my first and second replies because they do not think that bad faith or complicity undermine the moral standing to blame. I think that even if one is inclined to deny that bad faith or complicity are factors that undermine the moral standing to blame, then one should still accept that manipulation undermines the moral standing to blame. Whatever selection of replies one endorses, it should be clear that the MSMA does not undermine compatibilism.

Acknowledgements Thanks to an audience at the University of Manchester, including Helen Beebe, David Liggins, and Robert Knowles, for feedback on an earlier version of this paper. Thanks also to Patrick Todd and Michael McKenna for feedback. An extra thanks goes to Michael McKenna for encouraging me to continue with this paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Arpaly, N. (2006). Merit, meaning, and human bondage: An essay on free will. In *Princeton University press: Princeton*.
- Barnes, E. C. (2015). Freedom, creativity, and manipulation. *Noûs*, 49(3), 560–588.
- Capes, J. (2012). Blameworthiness without wrongdoing. *Pacific Philosophical Quarterly*, 93(3), 417–437.
- Cohen, G. A. (2006). Casting the first stone: Who can, and who Can't, condemn the terrorists? *Royal Institute of Philosophy Supplement*, 81(58), 113–136.
- Fischer, J. M. (2011). The zygote argument remixed. *Analysis*, 71(2), 267–272.
- Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press.
- Frankfurt, H. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy*, 68(1), 5–20.
- Frankfurt, H. (2002). Reply to John Martin Fischer. In *In Buss & Overton (eds.) Contour of Agency: Essays on Themes from Harry Frankfurt*. MIT press.
- Haji, I. (1998). *Moral Appraisability: Puzzles, proposals, and perplexities*. Oxford: Oxford University Press.
- Haji, I. (2013). Historicism, non-historicism, or a mix? *The Journal of Ethics*, 17(3), 185–204.
- Khouri, A. C. (2011). Blameworthiness and wrongness. *Journal of Value Inquiry*, 45(2), 135–146.
- Khouri, A. (2013). Synchronic and diachronic responsibility. *Philosophical Studies*, 165(3), 735–752.
- Mason, M. (2003). Contempt as a moral attitude. *Ethics*, 113(2), 234–272.
- Matheson, B. (2014). Compatibilism and personal identity. *Philosophical Studies*, 170(2), 317–334.
- McKenna, M. (2004). Responsibility and globally manipulated agents. *Philosophical Topics*, 32, 169–192.
- McKenna, M. (2008). A hard-line reply to Pereboom's four-case manipulation argument. *Philosophy and Phenomenological Research*, 77(1), 142–159.

- McKenna, M. (2014). Resisting the manipulation argument: A hard-liner takes it on the chin. *Philosophy and Phenomenological Research*, 89(2), 467–484.
- Mele, A. (2006). *Free will and luck*. Oxford: Oxford University Press.
- Mele, A. (2008). Manipulation, compatibilism, and moral responsibility. *The Journal of Ethics*, 12(3/4), 263–286.
- Russel, P. (2010). Selective hard compatibilism. In Campbell, J., O'Rourke, M., & Silverstein, H. (eds.) *Action, Ethics and Responsibility: Topics in Contemporary Philosophy* (Vol. 7). MIT Press: London.
- Talbert, M. (2009). Implanted desires, self-formation and blame. *Journal of Ethics and Social Philosophy*, 3(2), 1–18.
- Todd, P. (2012). 'Manipulation and moral standing: An argument for incompatibilism'. *Philosophers' Imprint*, 12(7), 1–18.
- Walzer, M. (1973). The problem of dirty hands. *Philosophy & Public Affairs*, 2(2), 160–180.