# The Routledge Companion to Free Will

Edited by Kevin Timpe, Meghan Griffith, and Neil Levy

# THE ROUTLEDGE COMPANION TO FREE WILL

*Edited by Kevin Timpe, Meghan Griffith, and Neil Levy*

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2017
by Routledge
711 Third Avenue, New York, NY 10017

and by Routledge
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

*Routledge is an imprint of the Taylor & Francis Group, an informa business*



*Library of Congress Cataloging in Publication Data*
A catalog record for this book has been requested

ISBN: 978-1-138-79581-5 (hbk)
ISBN: 978-1-315-75820-6 (ebk)

Typeset in Goudy Oldstyle Std
by Sunrise Setting Ltd, Brixham, UK

# 14
# THE MANIPULATION ARGUMENT

*Kristin Mickelson*

The Manipulation Argument has recently taken center stage in the free-will debate, yet little else can be said of this newcomer that is uncontroversial. At present, even the most fundamental elements of the Manipulation Argument—its structure, conclusion, and target audience—are a matter of dispute. As such, we cannot begin, as we ideally would, with a simple and relatively uncontroversial overview of the argument. Instead, clarifying the debate over the basic structure and general conclusion of the Manipulation Argument will be our goal.

In most discussions, the Manipulation Argument is understood as a *formal template* for an argument; each instance of the template is a distinct *manipulation argument*. The details of individual manipulation arguments greatly vary, but each proceeds something like this:

> Imagine that mischievous neuroscientists have developed technology which allows them to covertly invoke any mental states they like in their chosen victim, 'Vic.' The neuroscientists have grown tired of Vic's wife, so they press a series of buttons which cause Vic to undergo a process of reasoning which ends with his decision to kill her. Since there is nothing standing in Vic's way, he carries out the plan. Now, a question: *Is Vic free and morally responsible for killing his wife?* It certainly seems not: although Vic's decision to kill his wife is the causal product of his own inner states, these states are ultimately under the causal control of the neuroscientists. As such, Vic seems no more free—and, so, no more morally responsible—than a marionette. But, assuming that the laws of nature are deterministic, it seems that we, too, are mere marionettes: each of us is bound by 'causal strings' to facts in the distant past over which we had no control. Being subject to deterministic causal laws, then, is no different than being subject to freedom-undermining manipulation.

Because manipulation arguments typically compare scenarios involving freedom-undermining manipulation to scenarios involving deterministic laws of nature (where a realist view of the laws is tacitly assumed), it seems that such arguments indicate that there is an antagonistic metaphysical relationship between free action and deterministic laws. As such, the Manipulation Argument was originally considered a new way—and perhaps the *best* way (Taylor 1963: 45; Pereboom 2001: 89; McKenna 2010: 440)—to argue for the incompatibilist view that deterministic laws *undermine* free will.

However, recent work suggests that classifying the Manipulation Argument as an argument about the incompatibility of free will and deterministic laws may obscure the logic and lessons of its best instances. While all manipulation arguments target the view that it is possible for a normal human to act freely in a deterministic universe, the question of whether a properly fleshed out manipulation argument tells us something more—like *why* it is impossible for a human to act freely in a deterministic universe—has given rise to substantive debates about how to best understand these arguments. For example, Kristin Mickelson (2015b) and Neil Levy (2011) deny that manipulation arguments teach us that deterministic natural laws pose a threat to free will; instead, they propose that manipulation arguments support the conclusion that free action is impossible *tout court* due to the insurmountable problem of constitutive luck (this position is described in 'The Explanation Step' below). On this 'constitutive luck' interpretation, manipulation arguments challenge every view that is committed to the metaphysical possibility of free action, including free-will libertarianism (discussed in Ekstrom [Chapter 6], Griffith [Chapter 7], McCann [Chapter 8], this volume). At the very least, then, the standard classification of the Manipulation Argument as an 'argument for incompatibilism' misleadingly implies that there is a consensus regarding the challenges that manipulation arguments pose and to whom.

The dispute over the proper conclusion of the Manipulation Argument stems largely from the disagreement over the answer to one key question: must a successful instance of the Manipulation Argument *explain* why no one can act freely in a deterministic universe—and, if so, what is this explanation? Competing answers to this question give rise to rival views about how best to summarize the Manipulation Argument as a formal template, how that template is best fleshed out, and what conclusion its best instance(s) support. This chapter provides an introduction to the Manipulation Argument through a discussion of the most fundamental disagreements about its formal structure.

## The Structure of Manipulation Arguments

Contemporary interest in manipulation arguments is largely a response to Derk Pereboom's Four-case Argument (1995, 2001, 2014), the revitalization of an earlier manipulation argument from Richard Taylor (1963: 45, 46). Each of these manipulation arguments has three discrete steps: (i) Counterexample; (ii) Generalization; and (iii) Explanation. Each step provides the foundation for the conclusions drawn in subsequent steps. Roughly, the initial Counterexample Step concludes that philosophers have *not yet identified* a set of necessary and jointly sufficient conditions for free action according to which it is possible for a normal human person to act freely in a deterministic universe; the Generalization Step concludes (minimally) that there is *in principle* no such set; and the Explanation Step provides a freedom-denying explanation for *why* this is so. Let us look at each step in turn, using Pereboom's Four-case Argument as our example.

## The Counterexample Step

The Four-case Argument begins with a story of Professor Plum, hereafter 'Plum1,' who is manipulated by neuroscientists to kill Ms. White:

**Case 1**

Professor Plum was created by neuroscientists, who can manipulate him directly through the use of radio-like technology, but he is as much like an ordinary

> human being as is possible, given this history. Suppose these neuroscientists "locally" manipulate him to undertake the process of reasoning by which his desires are brought about and modified—directly producing his every state from moment to moment.
>
> (Pereboom 2001: 113)

Pereboom expects the intuitive judgement of Case 1—at least among members of the dialectically appropriate target audience (Mele 2008; Pereboom 2008; McKenna 2014)—will be that Plum1 is not free or morally responsible for killing Ms. White. Let us call this a 'victim judgment.' Yet, claims Pereboom, Plum1 satisfies all *freedom-neutral* metaphysical conditions on basic agency as well as the epistemic conditions on moral responsibility (2001: 111). As such, if Plum1 lacks moral responsibility, it is presumably because he lacks *free will* (understood roughly as the set of metaphysical, as opposed to epistemic, control conditions on moral responsibility). For this reason, the Four-case Argument—and every manipulation argument—is generally considered an argument about free will, even though it trades in intuitions about moral responsibility.

Pereboom contends that Plum1 satisfies all of the necessary conditions for free action hitherto proposed by advocates of the view that it is metaphysically possible for a normal human living in a deterministic universe to perform a free action. In defense of this claim, Pereboom provides a review of the most prominent of these conditions and contends that Plum1 satisfies each: (i) constancy of character; (ii) lack of constraint by irresistible desire; (iii) proper conformity of first-order and second-order desires; (iv) the capacity to regulate one's behavior based upon a moderately reasons-responsive deliberation process; and (v) the capacity to understand and regulate one's behavior based on moral reasons (Pereboom 2001: 100–110; 2014: 75). Assuming that Plum1 satisfies each of these conditions and yet Plum1 is not free or morally responsible for killing Ms. White, Case 1 reveals that these proposed necessary conditions are not *jointly sufficient* for free agency. Pereboom's Counterexample Step concludes that, even according to the intuitive judgments of the target audience, philosophers have yet to provide a set of necessary and jointly sufficient conditions for free action that can be satisfied by an ordinary human living in a universe with deterministic laws.

In order to properly understand the conclusion of the Counterexample Step and the remainder of the Four-case Argument, it is essential to recognize that Pereboom states his argument in terms of "human beings." While manipulation arguments can be understood as a strategy for teasing out which properties are relevant to *our* acting freely, Pereboom's argument seems to rest upon a substantive (and controversial) assumption about the nature of normal human beings. Prior to laying out the Four-case Argument in *Living Without Free Will*, Pereboom explicitly argues against the view that humans are agent causes, roughly the sort of agents who have a primitive "causal power to choose without being determined by events beyond the agent's control, and without the choice being a truly random or partially random event" (2001: 55). Pereboom argues that such agent causation is not "compatible with the physical world's being governed by exceptionless physical laws" (2001: 85), regardless of whether those laws are deterministic or indeterministic. Rather, Pereboom continues, it seems that agent causes would have to "override" such laws in order to act freely, but there is no evidence humans can do this (2001: 86; for alternative views on agent causation, see Griffith: Chapter 7, this volume). So, while Pereboom is sympathetic to the view that "overriding" agent causation is metaphysically possible and that such agents could perform a free action, he presents his

Four-case Argument against the background assumption that humans are not such agent causes; we humans are *subject to*—unable to override, trump, break, change, or otherwise perform a miracle relative to—the laws of nature.

Abstracting from the details of the Four-case Argument, we may summarize the Counterexample Step in less anthropocentric terms. The Counterexample Step is designed to elicit the intuitive judgment—a victim judgment—that the victim in the manipulation story lacks freedom and moral responsibility. It is expected that this victim judgment will be sufficiently strong and clear that, in effect, those who have it must accept that it provides a data point which any viable theory of free will must accommodate. However, the victim purportedly satisfies all hitherto proposed analyses of *free will* according to which it is possible for someone (human or not) who is subject to the laws of nature to act freely in a universe with deterministic laws. Assuming that the victim in the story satisfies these purportedly necessary conditions for free action and yet the manipulation story elicits a victim judgment, it seems that the Counterexample Step challenges the adequacy of *all* hitherto proposed analyses of *free will* according to which it is possible for someone (human or not) who is subject to the laws of nature to act freely in a universe with deterministic laws.

Because much of the basic vocabulary of the free-will debate is ambiguous (Mickelson 2015a), let us introduce terminology that will allow us to achieve the level of precision required for our discussion. Say that *compossibilism* is the view that it is metaphysically possible for someone to perform a free action in a universe with deterministic laws, and *incompossibilism* is the negation of compossibilism (Mickelson 2015b). Then say *compossibilism** is the narrower view that it is metaphysically possible for someone who is subject to deterministic laws to perform a free action, and its negation is *incompossibilism**. Roughly, then, the Counterexample Step concludes that there is currently no adequate compossibilism*-friendly analysis of the concept *free will*.

A final point about the language of the Counterexample Step will illuminate the significance of drawing a distinction between compossibilism* and compossibilism. The Four-case Argument—like most manipulation arguments—is stated in terms of *causal determination* and *deterministic physical laws* rather than the thesis of *determinism*. This is significant. Determinism is, roughly, the thesis that due to the past and the laws of nature, there is exactly one physically possible future (cf. van Inwagen 1983: 65). Given this definition of 'determinism,' and a realist view of the laws (a standard assumption in this context), it follows from the assumption that determinism is true that nothing—not even God—can make the future diverge from the one that is fixed by the natural laws (cf. Sehon 2011). In other words, if determinism is true, then everything that exists is *subject to* the laws of nature. As such, it would be impossible for an overriding agent cause (of the sort that Pereboom describes) to act freely when determinism is true. However, it is an open question whether miracles relative to the laws of nature are metaphysically possible. Indeed, Pereboom accepts that it is possible for something to violate (override, trump, break) deterministic natural laws in his defense of overriding agent causation. Assuming that it is possible for an overriding agent cause to perform a free action in a universe with deterministic laws, compossibilism is true. That is, Pereboom is sympathetic to compossibilism, and he does not argue that Case 1 is a counterexample to this view. By describing Plum1 as roughly an "ordinary human being" against the background assumption that humans are subject to the laws of nature, Pereboom narrows the target of his Counterexample Step from all extant compossibilism-friendly analyses of *free will* to contemporary accounts of compossibilism*.

Four-case Argument against the background assumption that humans are not such agent causes; we humans are *subject to*—unable to override, trump, break, change, or otherwise perform a miracle relative to—the laws of nature.

Abstracting from the details of the Four-case Argument, we may summarize the Counterexample Step in less anthropocentric terms. The Counterexample Step is designed to elicit the intuitive judgment—a victim judgment—that the victim in the manipulation story lacks freedom and moral responsibility. It is expected that this victim judgment will be sufficiently strong and clear that, in effect, those who have it must accept that it provides a data point which any viable theory of free will must accommodate. However, the victim purportedly satisfies all hitherto proposed analyses of *free will* according to which it is possible for someone (human or not) who is subject to the laws of nature to act freely in a universe with deterministic laws. Assuming that the victim in the story satisfies these purportedly necessary conditions for free action and yet the manipulation story elicits a victim judgment, it seems that the Counterexample Step challenges the adequacy of *all* hitherto proposed analyses of *free will* according to which it is possible for someone (human or not) who is subject to the laws of nature to act freely in a universe with deterministic laws.

Because much of the basic vocabulary of the free-will debate is ambiguous (Mickelson 2015a), let us introduce terminology that will allow us to achieve the level of precision required for our discussion. Say that *compossibilism* is the view that it is metaphysically possible for someone to perform a free action in a universe with deterministic laws, and *incompossibilism* is the negation of compossibilism (Mickelson 2015b). Then say *compossibilism** is the narrower view that it is metaphysically possible for someone who is subject to deterministic laws to perform a free action, and its negation is *incompossibilism**. Roughly, then, the Counterexample Step concludes that there is currently no adequate compossibilism*-friendly analysis of the concept *free will*.

A final point about the language of the Counterexample Step will illuminate the significance of drawing a distinction between compossibilism* and compossibilism. The Four-case Argument—like most manipulation arguments—is stated in terms of *causal determination* and *deterministic physical laws* rather than the thesis of *determinism*. This is significant. Determinism is, roughly, the thesis that due to the past and the laws of nature, there is exactly one physically possible future (cf. van Inwagen 1983: 65). Given this definition of 'determinism,' and a realist view of the laws (a standard assumption in this context), it follows from the assumption that determinism is true that nothing—not even God—can make the future diverge from the one that is fixed by the natural laws (cf. Sehon 2011). In other words, if determinism is true, then everything that exists is *subject to* the laws of nature. As such, it would be impossible for an overriding agent cause (of the sort that Pereboom describes) to act freely when determinism is true. However, it is an open question whether miracles relative to the laws of nature are metaphysically possible. Indeed, Pereboom accepts that it is possible for something to violate (override, trump, break) deterministic natural laws in his defense of overriding agent causation. Assuming that it is possible for an overriding agent cause to perform a free action in a universe with deterministic laws, compossibilism is true. That is, Pereboom is sympathetic to compossibilism, and he does not argue that Case 1 is a counterexample to this view. By describing Plum1 as roughly an "ordinary human being" against the background assumption that humans are subject to the laws of nature, Pereboom narrows the target of his Counterexample Step from all extant compossibilism-friendly analyses of *free will* to contemporary accounts of compossibilism*.

## The Generalization Step

In the Generalization Step of the Four-case Argument, Pereboom uses the victim judgment elicited in the Counterexample Step to motivate a generalization argument. This generalization argument is (following McKenna 2004) most commonly summarized as follows:

1. *Victim Premise*. Plum1 lacks freedom and moral responsibility for killing Ms. White.
2. *No-difference Premise*. There is no freedom-relevant or responsibility-relevant difference between Plum1's killing Ms. White and any action performed by a normal human living in a universe at which the laws of nature are deterministic (i.e., a deterministic universe).
3. *Conclusion*. So, no normal human living in a deterministic universe ever performs a free action.

The sole support for the Victim Premise is the victim judgment elicited in the Counterexample Step. So, assuming that the initial Counterexample Step is successful, the proponent of a manipulation argument need not offer any *additional* support for the Victim Premise.

The main support for the No-difference Premise comes in the form of a *No-difference Defense*. This No-difference Defense proceeds by transforming the initial covert manipulation case "in perfectly realistic ways, so as to coincide with actual and familiar cases" (Taylor 1963: 45, 46). This procedure results in a series of 'bridge' cases that span between a manipulation case on one end and a 'normal' case on the other. Each new bridge case is created by removing some apparently freedom-relevant feature *F* that is present in previous cases of the series. The new case (without *F*) is then compared to early cases in the series, and it is argued that there is no freedom-relevant difference between the new case and the original manipulation case. The conclusion is then drawn that *F* is not a freedom-relevant difference after all. The goal of this process is to eliminate all of the apparent freedom-relevant differences between the manipulation scenario and the normal scenario, thereby supporting the claim that the protagonist in each scenario of the series has the same status with respect to free agency.

Pereboom's No-difference Defense of Premise 2 begins with bridge cases Case 2 and Case 3. The actions of Plum in Case 2, hereafter 'Plum2,' are *indirectly* controlled by the neuroscientists through covert programing:

> Plum is like an ordinary human being, except that he was created by neuroscientists who, although they cannot control him directly, have programmed him to weigh reasons for action so that he is often but not exclusively egoistic, with the result that in the circumstances in which he now finds himself, he is causally determined to undertake the [process of reasoning that results in his killing Ms. White].
>
> (Pereboom 2001: 113, 114)

Pereboom expects that Case 2 will elicit the same intuitive reaction as Case 1—namely, the victim judgment that Plum2 does not act freely or morally responsibly when he kills Ms. White. Indeed, Pereboom proposes that Case 2 could serve as the foundational counterexample story should Case 1 fail—say, for example, because Plum1 fails

the necessary conditions of *agency* (Fischer and Ravizza 1998: 234, 235, n. 28; Mele 2005: 78; Baker 2006: 320; Demetriou 2010). According to Pereboom, Case 2 allows us to see that eliminating the *moment-by-moment causal control* exerted by the neuroscientists in Case 1 does not transform Plum into a free and responsible agent, so this feature of Case 1 is not freedom-undermining. Next, Pereboom describes Case 3, a near-normal situation in which overbearing parents impose rigorous training on young Plum. Finally, Pereboom argues there is no principled difference between the Plums in the first three cases and Case 4, a case in which Plum ('Plum4') is a perfectly normal human being who is born into a deterministic universe. As such, the No-difference Defense shows that anyone who denies that Plum1 is a free and responsible agent must, on the pain of inconsistency, make the same judgment of Plum4.

There is no magical number of bridge cases that a No-difference Defense must have, and new bridge cases may be added as needed to meet new objections. For instance, after laying out his initial four cases, Pereboom offers a few more—which favors the Four-case Argument's alternative name, the "Multiple-case Argument" (Pereboom 2008). Among others, Pereboom forwards a case in which Plum's states are spontaneously induced by a machine rather than by external agents. This bridge case is designed to forestall the specific proposal that *having one's states induced by another agent or intelligent designer* is a freedom-relevant feature difference between Case 1 and Case 4 (Pereboom 2001: 114–16). Additional cases and supplemental argument might be added to address any other proposed freedom-relevant difference between Plum1 and Plum4 (e.g., McKenna 2008: 152, 153). If Pereboom is right that there is no freedom- or responsibility-relevant difference between these Plums, then either all of the Plums are free and morally responsible for killing Ms. White or none of them is; the No-difference Premise is true.

The conclusion of Pereboom's generalization argument is restricted to humans who are living in a deterministic universe, where humans are assumed to be among the class of metaphysically possible beings who are subject to the causal laws. However, Pereboom's proposed generalization argument technically leaves open the question of whether *being subject to deterministic laws* and/or *living in a deterministic universe* are freedom-relevant features of the cases. In principle, Pereboom's No-difference Defense could be expanded to eliminate the qualifications that appear in his preferred statement of the No-difference Premise. This expanded argument would result in a stronger instance of the No-difference Premise and the resulting manipulation argument would reach a bolder conclusion. For example, working along such lines, Mele forwards Case 2a (as part of a critique of the Four-case Argument) in which Plum2's deterministic programming is replaced with an indeterministic version:

> It [the program in Case 2a] works just like the program in case 2 except that there is a tiny chance every few seconds that the program will incapacitate Plum. As it happens, Plum is not incapacitated. If Plum is not morally responsible for the killing in case 2, he is not morally responsible for it in case 2a either. Surely, blending this possibility of incapacitation into case 2 does not transform it from a case of non-responsibility into one of responsibility. Here again that the causation in Pereboom's case is deterministic is not essential to Plum's lacking moral responsibility for the killing.
>
> (Mele 2005: 76)

Mele proposes that, intuitively, adding indeterministic causation into the scenario does not "transform" an unfree agent into a free one. Assuming that Mele is right—*pace* event-causal libertarians such as Robert Kane (1996)—that adding indeterministic causation does not *help* Plum to have free will, Case 2a casts doubt on Pereboom's proposal that deterministic causation in the original scenario positively *hurts*. The addition of Case 2a to Pereboom's original series of cases shows how one might develop a No-difference Defense that rules out the deterministic causation as a freedom-undermining feature in Pereboom's original four cases. If deterministic causation is not a freedom-relevant feature of the cases, then Pereboom's inclusion of this constraint *arbitrarily* limits the No-difference Premise's scope and, by extension, the ultimate conclusion of the Generalization Step. This means that the 'lives in a deterministic universe' constraint can, and perhaps should, be eliminated from both Pereboom's No-difference Premise and the conclusion of his generalization argument.

Assuming that deterministic laws (*qua* being deterministic) do not pose a threat to free will, perhaps the problem is *being subject to causal laws* (irrespective of whether those laws are deterministic or probabilistic). A manipulation argument could be developed to test this proposal. For example, one might offer a bridge case describing an actor—such as Pereboom's law-trumping agent cause—who is *not* subject to the laws of nature. Using this case, one could develop a No-difference Defense of the conclusion that (*pace* Pereboom) *being subject to the causal laws* is not a freedom-relevant property. Continuing along such lines, one might argue that there is no freedom-relevant difference between the manipulation victim and someone who satisfies the necessary and sufficient conditions of being an agent cause (King 2013: 72).

In principle, a No-difference Defense might be used to address every purportedly freedom-relevant feature of the original manipulation case, thereby eliminating all constraints from Pereboom's original No-difference Premise. The resulting No-difference Premise would state that there is no freedom-relevant difference between the manipulation victim and *any metaphysically possible being* (including God) in *any metaphysically possible conditions* (e.g., whether the laws are indeterministic or non-existent). By extension, the resulting generalization argument would be an argument for unqualified free-will *impossibilism*, the view that free action is metaphysically impossible. Such a manipulation argument would constitute a challenge to all compossibilist and libertarian views alike.

In summary, the conclusion of the Generalization Step may vary widely, depending on how the No-difference Defense is fleshed out. However, this flexibility in the structure and conclusion of manipulation arguments is not captured by the standard formal summary of the Pereboom's generalization argument. It seems, then, that Pereboom's proposed generalization argument is better understood as an instance of a more generic template that proceeds something like this:

## The Generalization Argument Template

1. *Victim Premise*. The victim *V* (in the initial manipulation case) is not free or responsible for performing action *A*.
2. *Generalization Premise*. If (and only if) *V* is not free or responsible for performing *A*, then no *B-type being* living in *C-type conditions* performs a free action for which he is also morally responsible.
3. *Conclusion*. No *B-type being* living in *C-type conditions* performs a free action for which he is also morally responsible.

Notably, the Generalization Argument Template (hereafter, the 'Generalization Argument') does not have a 'No-difference' premise. This is because the biconditional Generalization Premise more explicitly captures the key inference of the Generalization Step. An instance of the Generalization Premise is, minimally, supported by a No-difference Defense and the assumption that like cases should be treated alike (McKenna 2012: 150). Since every instance of the Generalization Argument is valid, the only remaining question is whether, given the details of a specific manipulation story, any instance has true premises.

Given that the details of the foundational manipulation story are critical to the success of both the Counterexample and Generalization Steps, it is worth noting that some prominent manipulation cases have limited potential when understood as manipulation arguments. For instance, Mele's Ann/Beth cases (Mele 1995: 145, 146, 2006: 164–7) and McKenna's Suzie Instant/Suzie Normal cases (McKenna 2004: 180, 181, 2012: 160–6) are primarily used to mediate an in-house debate among *historical* and *ahistorical* compossibilists*. The victims in these manipulation stories do not satisfy any history-sensitive conditions on free will, for example, Mele's "bypass" condition (Mele 1995). However, they do seem to satisfy all hitherto proposed *ahistorical* necessary conditions on free action, such as Harry Frankfurt's requirement of a proper alignment between one's first-order and second-order desires (Frankfurt 1971). As such, these manipulation cases may constitute counterexamples to extant *ahistorical* analyses of free will, but not to their *history-sensitive* rivals. So, while there is no in-principle reason that such cases cannot be used in a manipulation argument (McKenna 2012: 169), proponents of history-sensitive compossibilism* are well-positioned to reject any such manipulation argument both on the grounds that such stories do not constitute counterexamples to their preferred history-sensitive analyses of free will and that there is a freedom-relevant difference that blocks the generalization from the manipulation scenario to normal deterministic scenarios.

## The Explanation Step

Most formal summaries of the Four-case Argument characterize it as a mere combination of the Counterexample and Generalization Steps, but these two steps do not fully capture Pereboom's original argument. Pereboom—like Taylor before him (1963: 46)—argues that the protagonist in each of his cases lacks freedom and responsibility *because* the decision to kill is an "alien-deterministic event," that is, a decision "produced by a deterministic process that traces back to causal factors beyond [the actor's] control" (Pereboom 2001: 126). As Pereboom develops his manipulation argument, then, it is an argument for *incompatibilism*, understood roughly as the thesis that necessarily, if someone is subject to deterministic causal laws then that person lacks free will (at least in part) *because* she is subject to deterministic laws (McKenna 2010: 432; Levy 2011; Mickelson 2015a). Indeed, Pereboom contends that the Four-case Argument concludes specifically to *causal-history* or *source* incompatibilism, an explanatory subtype of incompatibilism according to which causal determination is a threat to free will *because* it prevents one from being the freedom-relevant source of his own actions (Pereboom 1995; McKenna 2000; King 2013: 79). However, the logical form of the Generalization Argument guarantees that every instance has a negative non-explanatory conclusion, such as incompossibilism* or impossibilism; no instance concludes with the positive identification of a freedom-undermining feature that is present in both the manipulation case and the end case. Since the Counterexample and Generalization Steps alone

cannot capture the conclusion of the Four-case Argument, Pereboom's argument is not adequately represented by these steps alone.

Pereboom upgrades his Four-case Argument from an argument for mere incompossibilism* to an argument for incompatibilism by adding a best-explanation argument. According to Pereboom the *best explanation* for Plum's lack of free agency in Case 1 through Case 4 is that the Plum in each case is subject to deterministic causation and this prevents him from being the freedom-relevant source of his actions laws (2001: 112–15). Pereboom sets the stage for this best-explanation argument in the Generalization Step, where he uses a No-difference Defense to rule out anticipated but (purportedly) unviable explanations for Plum1's lack of free agency. However, it is not until the Explanation Step of the Four-case Argument that Pereboom finally completes the positive task of identifying *deterministic causation* as the freedom-undermining feature that is present in both the initial manipulation case and the normal end case.

Although Pereboom develops the Four-case Argument as an argument for source incompatibilism and manipulation arguments are commonly thought to be homing in on sourcehood requirements for free action, there is growing disagreement about whether manipulation arguments favor a specifically *incompatibilist* notion of sourcehood. As noted above, it is in principle possible to develop an instance of the Generalization Argument that concludes to impossibilism. Assuming that the goal of the Explanation Step is to provide the best explanation for the impossibility of free will, one might argue—as some have (Levy 2011: 86–9; King 2013: 78; Mickelson 2015b)—that the lack of free action in the manipulation and normal scenarios is due to a pernicious sourcehood problem known as *constitutive luck*. Following Thomas Nagel (1979), Levy describes constitutive luck as "luck in the traits and dispositions that make one the kind of person one is" (Levy 2011: 29). With his "Basic Argument," Galen Strawson famously argues that constitutive luck poses an insurmountable threat to free will (cf. Strawson 1986: 28–9). According to Strawson, free action requires "a starting point in the series of acts of [intentionally] bringing it about that one has a certain nature—a starting point that constitutes an act of ultimate self-origination" (Strawson 2011). Only a *causa sui* (a being who self-creates *ex nihilo*) would satisfy Strawson's stringent starting-point condition; anyone who does not satisfy this starting-point condition suffers from freedom-undermining constitutive luck. However, assuming that such robust self-creation is metaphysically impossible, it follows that there is no possible agent who ever satisfies this starting-point condition. Whether the laws are deterministic or indeterministic is irrelevant to the fact that no actor is a *causa sui* at any given time at which that actor exists, so the constitutive-luck explanation seems to imply that the laws of nature are totally irrelevant to the fact that no one ever performs a free action. This demonstrates that proponents of the constitutive-luck explanation may agree with Pereboom that each of the Plums has a freedom-undermining sourcehood problem and yet deny that it is metaphysically possible for deterministic laws to play a role in *undermining* free will (even in the modest sense that such laws prevent people from *overcoming* constitutive luck). In other words, if the constitutive-luck diagnosis is correct, then impossibilism is true, but incompatibilism—in virtue of its mistaken explanatory thesis—is false.

Whatever specific freedom-undermining feature is ultimately identified in the Explanation Step, the proposed diagnosis seems to clarify and reaffirm the conclusions drawn in earlier steps of the manipulation argument in at least three ways. First, the diagnosis forwarded in the Explanation Step of a manipulation argument illuminates the *scope* of

the conclusion of the Generalization Step. As discussed above, there is no in-principle reason that an instance of the Generalization Argument must conclude to a restricted thesis such as incompossibilism* rather than to incompossibilism or the completely unrestricted thesis of impossibilism. As such, any limitations on the scope of an instance of the Generalization Premise and, by extension, the conclusion of an instance of the Generalization Argument seem arbitrary in the absence of a diagnosis of the freedom-undermining feature that is common to all cases. For example, only by proposing that deterministic laws undermine the free will of anyone who is subject to them does Pereboom illuminate why his generalization argument concludes to incompossibilism* rather than to some less restricted thesis. Second, a best-explanation argument adds *positive* support for a proposed instance of the Generalization Premise by identifying the freedom-undermining feature that is (purportedly) present in both the initial manipulation case and normal end case. Finally, it seems that the Explanation Step may bolster the Counterexample Step. The central purpose of the Explanation Step is to identify the specific freedom-undermining feature *F* that is present in each of the cases, from the initial manipulation case to the normal end case. As such, the proposed explanation for the victims' lack of freedom and responsibility suggests a second best-explanation argument, this one aimed at explaining the intuitive judgement that the manipulation victim lacks free will: the best explanation for one's victim judgment is that it is a response to *F*. As such, the proposed explanation seems to sanction a key methodological background assumption of the Counterexample Step, namely that the victim judgment is a *rational* response to some freedom-undermining feature of the initial manipulation case —an assumption that has not been immune to criticism (e.g., McKenna 2008: 157; Spitzley 2015). Of course, it is ultimately an empirical question why a particular manipulation case elicits a particular intuition from a particular person, and it is far from obvious that every feature tracked by a victim judgment (even if that judgment is correct and its source can be identified) is *metaphysically relevant* to free will.

## Non-diagnostic Manipulation Arguments

While the Explanation Step provides a diagnosis that both answers pressing philosophical questions and nicely rounds out the overall manipulation argument, not everyone agrees that manipulation arguments *require* an Explanation Step (Mele 2008: 278; Pereboom 2014: 79, 80, n. 3; Mickelson 2015b). Since the purpose of the Explanation Step is to give a positive diagnosis of someone's lack of freedom and moral responsibility, let us classify manipulation arguments that flesh out the Explanation Step as 'diagnostic' arguments and those that do not as 'non-diagnostic.' A closer look at Alfred Mele's Zygote Argument, the most prominent non-diagnostic manipulation argument in the current literature, will help us to decide whether the benefits of dropping the Explanation Step are outweighed by the costs.

Mele begins the Zygote Argument with a story that is, in its basic metaphysical details, similar to Pereboom's Case 2:

### The Zygote Case

> Diana creates a zygote Z in Mary. She combines Z's atoms as she does because she wants a certain event E to occur 30 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 30 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. If this agent, Ernie, has any unsheddable values at the time, they play no role in motivating his A-ing. Thirty years later, Ernie is a mentally healthy, ideally self-controlled person who regularly exercises his powers of self-control and has no relevant compelled or coercively produced attitudes. Furthermore, his beliefs are conducive to informed deliberation about all matters that concern him, and he is a reliable deliberator.
>
> (Mele 2006: 188, 2013: 175, 176)

Mele proceeds under the assumption that the Zygote Story will elicit (in at least some members of its target audience) the intuition that Ernie lacks freedom and moral responsibility for *A-ing*. This intuition provides support for Premise 1 of the Zygote Argument, which proceeds as follows:

**ZAM**

1. Ernie is not a free agent and is not morally responsible for anything.
2. Concerning free action and moral responsibility of the beings into whom the zygotes develop, there is no significant difference between the way Ernie's zygote comes to exist and the way any normal human zygote comes to exist in a deterministic universe.
3. So in no possible deterministic world in which a *human being develops from a normal human zygote* is that human being morally responsible for anything he or she does.

(Mele 2013: 176, my emphasis)

With this generalization argument, Mele's version of the Zygote Argument comes to an end; there is no crowning Explanation Step.

Mele recognizes that the Zygote Argument and the Four-case Argument have a different formal structure. According to Mele, premise 2 of ZAM is a negative "no-difference" claim while the second premise of the Four-case Argument is a positive best-explanation claim (2008). However, Mele describes both of these manipulation arguments as ending with roughly the same conclusion. Overall, Mele's description gives the impression that each of these arguments is best understood as an instance of the Generalization Argument, where the main formal difference between the two is the way they support their respective generalization premises: the Four-case Argument includes positive support in the form of a best-explanation argument, but the Zygote Argument does not. So framed, the best-explanation argument in the Four-case Argument seems superfluous, for a No-difference Defense provides adequate support for the key generalization inference. This framing leads Mele to suggest that one "fallback" position open to proponents of the Four-case Argument is to drop its best-explanation argument (Mele 2008: 278). The main benefit of adopting this fallback position is that the resulting version of the Four-case Argument would be immune to criticisms targeting its best-explanation argument (Mele 2008: 276–8). However, the proposed fallback position also comes at a cost: to drop the best-explanation argument is to forsake the Four-Case Argument's original, explanatory conclusion that deterministic causation poses a threat to free will.

To make the cost of dropping the Explanation Step still clearer, consider the difference between the conclusion of ZAM and the conclusion of the Generalization Step of the Four-case Argument: the Four-case Argument concludes to the modest thesis of incompossibilism*, but ZAM's final conclusion is more modest still. ZAM's conclusion is restricted to the narrow subset of metaphysically possible beings that are *human*, *developed from a normal human zygote*, and who *live in a universe with deterministic laws*. But is the property of *being human* or *having developed from a normal human zygote* a freedom-relevant feature of a person? Or, more generally, is having *a first moment of existence* relevant to free action (Campbell 2007; Bailey 2012)? How about whether an agent is *able to break or override causal laws*? Do *deterministic laws* preclude freedom-relevant sourcehood—or is the only genuine threat to free action *constitutive luck*? A non-diagnostic manipulation argument such as ZAM, does not—and cannot—answer any of these historically popular and philosophically pressing questions.

## Conclusion

Summing up, there is presently no uncontentious formal characterization of the Manipulation Argument. Minimally, it seems that the Manipulation Argument outlines a persuasive argument for a version of the non-explanatory thesis that necessarily, no one who is subject to the laws can act freely in a deterministic universe. However, it remains unsettled whether the Manipulation Argument is also essentially an argument for some positive, explanatory thesis—and, if so, whether this thesis is incompatibilism, as the first manipulation arguments suggest, or some other explanatory view, as the newer constitutive-luck variants suggest. Still, the Manipulation Argument has been instrumental in exposing the explanatory gap between views such as incompossibilism and incompatibilism, and it provides an excellent framework for future discussions on whether and how best to close it. Such contributions make the Manipulation Argument worthy of its pride of place in the contemporary free-will debate and indicate that the argument will continue to bear fruit for many years to come.

## Bibliography

Bailey, A. (2012) "Incompatibilism and the Past," *Philosophy and Phenomenological Research* 85: 351–76.
Baker, L. (2006) "Moral Responsibility without Libertarianism," *Noûs* 40: 307–30.
Campbell, J. (2007) "Free Will and the Necessity of the Past," *Analysis* 67: 105–11.
Demetriou, K. (see also K. Mickelson) (2010) "The Soft-Line Solution to Pereboom's Four-case Argument," *Australasian Journal of Philosophy* 88: 595–617.
Fischer, J. and 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: 5–20.
Kane, R. (1996) *The Significance of Free Will*. Oxford: Oxford University Press.
King, M. (2013) "The Problem with Manipulation," *Ethics* 124: 65–83.
Levy, N. (2011) *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. New York: Oxford University Press.
McKenna, M. (2000) "Source Incompatibilism, Ultimacy, and the Transfer of Non-responsibility," *American Philosophical Quarterly* 38: 37–51.
McKenna, M. (2004) "Responsibility and Globally Manipulated Agents," *Philosophical Topics* 32: 169–92.
McKenna, M. (2008) "A Hard-Line Reply to Pereboom's Four-Case Argument," *Philosophy and Phenomenological Research* 77: 142–59.
McKenna, M. (2010) "Whose Argumentative Burden, Which Incompatibilist Arguments? – Getting the Dialectic Right," *Australasian Journal of Philosophy* 88: 429–43.

McKenna, M. (2012) "Moral Responsibility, Manipulation Arguments, and History: Assessing the Resilience of Nonhistorical Compatibilism," *Journal of Ethics* 16: 145–74.
McKenna, M. (2014) "Resisting the Manipulation Argument: A Hard-Liner Takes It on the Chin," *Philosophy and Phenomenological Research* 89: 467–84.
Mele, A. (1995) *Autonomous Agents*. New York: Oxford University Press.
Mele, A. (2005) "A Critique of Pereboom's 'Four-Case Argument' for Incompatibilism," *Analysis* 65: 75–80.
Mele, A. (2006) *Free Will and Luck*. New York: Oxford University Press.
Mele, A. (2008) "Manipulation, and moral responsibility," *Journal of Ethics* 12: 263–86.
Mele, A. (2013) "Manipulation, Moral Responsibility, and Bullet Biting," *Journal of Ethics* 17: 167–84.
Mickelson, K. (2015a) "A critique of Vihvelin's Threefold Classification," *Canadian Journal of Philosophy*, 45: 85–99.
Mickelson, K. (see also K. Demetriou) (2015b) "The Zygote Argument is Invalid–Now What?" *Philosophical Studies* 172: 2911–29.
Nagel, T. (1979) "Moral luck," in *Mortal Questions*, New York: Cambridge University Press, pp. 24–38.
Pereboom, D. (1995) "Determinism al Dente," *Noûs* 29: 21–45.
Pereboom, D. (2001) *Living Without Free Will*. Cambridge: Cambridge University Press.
Pereboom, D. (2008) "A Hard-Line Reply to the Multiple-Case Manipulation Argument," *Philosophy and Phenomenological Research* 77: 160–70.
Pereboom, D. (2014) *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.
Sehon, S. (2011) "A Flawed Conception of Determinism in the Consequence Argument," *Analysis* 71: 30–8.
Spitzley, J. (2015) "The Importance of Correctly Explaining Intuitions: Why Pereboom's Four-Case Manipulation Argument is Manipulative," *Journal of Cognition and Neuroethics* 3(1): 363–82.
Sripada, C. (2012) "What Makes a Manipulated Agent Unfree?" *Philosophy and Phenomenological Research* 85: 563–93.
Strawson, G. (1986) *Freedom and Belief*. Oxford: Clarendon Press.
Strawson, G. (2011) "Free Will" in E. Craig (ed.), *Routledge Encyclopedia of Philosophy*, London: Routledge, available from: www.rep.routledge.com/articles/free-will/v-2/, DOI:10.4324/9780415249126-V014-2 (accessed 10 July 2016).
Taylor, R. (1963) *Metaphysics*. Engelwood Cliffs: Prentice-Hall, Inc.
van Inwagen, P. (1983) *An Essay on Free Will*. New York: Oxford University Press/Clarendon Press.

## Further Reading

For an excellent overview of the literature, see D. Pereboom's "A Manipulation Argument against Compatibilism," in *Free Will, Agency, and Meaning in Life* (2014: Chapter 4). Replies to the Manipulation Argument are often classified based on the premises of the Generalization Argument: a 'hard-line' reply rejects the Victim Premise while a 'soft-line' reply rejects the Generalization Premise. For an introduction to such replies, see M. McKenna (2012, 2014); D. Pereboom (2008); Fischer, J.M. (2011) "The Zygote Argument remixed," *Analysis* 71: 267–72; K. Demetriou (2010); Haji, I. and Cuypers, S. (2006) "Hard- and Soft-Line Responses to Pereboom's Four-Case Manipulation Argument," *Acta Analytica* 21: 19–35. For replies targeting the Explanation Step, see Mickelson (2015b) and Mele (2005). For alternative response strategies, see Kearns, S. (2012) "Aborting the Zygote Argument," *Philosophical Studies* 160: 379–89; Todd, P. (2011) "A New Approach to Manipulation Arguments," *Philosophical Studies* 152: 127–33; King (2013); Tognazzini, N. (2014) "The Structure of a Manipulation Argument," *Ethics* 124: 358–69. For a synopsis of relevant experimental philosophy, see (esp. Section 9.4) of G. Björnsson and D. Pereboom's "Traditional and Experimental Approaches to Free Will and Moral Responsibility," in J. Sytsma and W. Buckwalter (eds), *A Companion to Experimental Philosophy* (2016).

## Related Topics

Skeptical Views about Free Will
The Consequence Argument
Leeway vs. Sourcehood Conceptions of Free Will
Folk Intuitions
Free Will and Theological Fatalism
Free Will and Providence
Determinism