**FREEDOM, MORAL RESPONSIBILITY, AND THE FAILURE OF UNIVERSAL DEFEAT**

Andrew J. Latham

Aarhus University

Somogy Varga

Aarhus University

Hannah Tierney

University of California, Davis

1. **Introduction**

Compare the following two cases:

Case 1: Katie is essentially a normal woman, but unlike everyone else, aliens implanted a device in her brain at the moment of her birth that causally determines Katie to make decisions that almost always benefit her. The aliens did this because they wanted Katie to one day kill Mr. Plum and they knew that Katie would do so if she was determined to make decisions that almost always benefited her. Katie could not have prevented the aliens from causing her to make decisions this way and it is ingrained in her. One day, due to being causally determined to make decisions that almost always benefit her, Katie decides to kill Mr. Plum.

Case 2: Katie\* is a normal woman, and like everyone else, aliens implanted a device in her brain at the moment of her birth that causally determine Katie\* to make decisions that almost always benefit her. The aliens did this because they wanted Katie to one day kill Mr. Plum and they knew that Katie would do so if she was determined to make decisions that almost always benefited her. Katie\*, just like everyone else, could not have prevented the aliens from causing her to make decisions this way and it is ingrained in her. One day, due to being causally determined to make decisions that almost always benefit her, Katie\* decides to kill Mr. Plum.

Notice the only difference between Case 1 and Case 2 is the scope of the manipulation. Katie is the *only* agent in her world that is being manipulated by an implanted alien device. In contrast, Katie\* is just like *everyone* else; everyone in her world is being manipulated by implanted alien devices. Following Andrew James Latham and Hannah Tierney (2021, 2022), we will call a manipulation whose scope includes at least one but not all agents an **existential manipulation** and a manipulation whose scope includes every agent a **universal manipulation**. Is the difference in the scope of a manipulation an *important* one when it comes to free will and moral responsibility? That is, does a manipulation being existential or universal affect the degree to which we judge Katie and Katie\* to be free and morally responsible?

 Intuitively, one might think that the answer should be *no*. If we judge that some factor undermines the free will and moral responsibility of one agent, then it stands to reason that we would take this factor to undermine the free will and responsibility of any agent who is impacted by this factor, even if all agents are so impacted. So, if we thought that Katie in Case 1 was not free and morally responsible with respect to her decision to kill Mr. Plum due to being manipulated by an implanted alien device, then we should also think that Katie\* in Case 2, along with everyone else in her world, is not free and responsible as well. Or so says current orthodoxy. The assumption that the scope of manipulation is neither important nor relevant to free will and moral responsibility plays an important role in manipulation arguments, an influential style of argument against compatibilism.[[1]](#endnote-1) These arguments typically proceed by securing the intuition that manipulated agents are not free and responsible, usually by using cases of existential manipulation (i.e., Katie being manipulated by an implanted alien device). Defenders of these arguments then argue that existential manipulation is relevantly similar to causal determinism—the truth of causal determinism rules out free will and moral responsibility in the same way that existential manipulation does. However, determinism is a universal thesis—if true, all agents, and all things, fall within its scope. Thus, these arguments can only work if there really is no relevant difference between existential manipulation and universal theses like determinism.

 While the assumption that **manipulation scope** is irrelevant to our free will and moral responsibility judgments has gone largely unquestioned in the literature, the claim that there are *no* important and relevant differences between manipulation and determinism cases is highly contested. Many opponents of manipulation arguments have argued that **manipulation type** is relevant to these judgments. To illustrate, consider another case, this time involving Katie\*\*.

Case 3: Katie\*\* is a normal woman, and like everyone else, the past facts and laws of nature causally determine Katie\*\* to make decisions that almost always benefit her – that is, given the past and laws of nature, everyone is determined to make decisions that almost always benefit them, including Katie\*\*. Katie\*\*, just like everyone else, could not have prevented being determined to make decisions this way and it is ingrained in her. One day, due to being causally determined to make decisions that almost always benefit her, Katie\*\* decides to kill Mr. Plum.

Comparing Case 2 to Case 3, there is no difference in scope: Both Katie\* and Katie\*\* are targets of universal manipulation.[[2]](#endnote-2) The only difference between Cases 2 and 3 is manipulation type:[[3]](#endnote-3) Katie\* is causally determined by an implanted alien device, whereas Katie\*\* is causally determined by the laws of nature in conjunction with the past facts. For many opponents of manipulation arguments, this represents an important and relevant difference. According to these theorists, Katie\* is not free and morally responsible with respect to her decision to kill Mr. Plum because she is causally determined to do so by *other agents who intend for her to decide to kill Mr. Plum*.[[4]](#endnote-4) In contrast, Katie\*\* is free and responsible because, while she is causally determined to decide to kill Mr. Plum, it is not because of the influence of other agents.

 However, recent work by Latham and Tierney (2021, 2022) poses a challenge to both proponents of manipulation arguments and their opponents. In one study, they found evidence that people’s free will and moral responsibility judgments are sensitive to what they call the existential/universal effect: people judge agents impacted by existential manipulation to be significantly less free and responsible than agents impacted by universal manipulation. Conversely, they found *no* evidence that manipulation type had any effect on people’s free will and moral responsibility judgments (2021: 2). These findings indicate that there may be a difference between existential and universal forms of manipulation that has gone unnoticed by defenders of manipulation arguments and their opponents. Perhaps manipulation arguments fail because free will and moral responsibility judgments are sensitive to manipulation scope, rather than manipulation type.

 While Latham and Tierney’s research is suggestive, their findings are limited. For example, their manipulation cases feature agents whose decisions are determined by their upbringing, which doesn’t accentuate the role of manipulative agents as much as other kinds of manipulation (e.g., alien interventions). Moreover, their cases don’t clearly involve forms of *intentional* manipulation, where manipulators *intend* that an agent decides to perform a particular action and then intervene to ensure that they do so. According to some critics of manipulation arguments (e.g., Deery & Nahmias 2017; Waller 2014), the intention with which manipulators act is what sets manipulation apart from determinism, and Latham and Tierney’s research does not bear on this hypothesis.

 In this paper, we build on Latham and Tierney’s research and investigate a wider range of cases featuring different forms of manipulation and determinism. Specifically, we examined people’s judgements about existential and universal manipulation cases that clearly feature manipulative agents and involve both intentional and non-intentional outcomes. We show for the first time that people’s judgments in manipulation cases involving *intentional* outcomes are sensitive to the existential/universal effect. People judged existentially manipulated agents with an intentional outcome to be significantly less free and responsible than universally manipulated agents with an intentional outcome. Thus, we replicate the existential/universal effect, confirming that manipulation scope impacts judgments about free will and moral responsibility, indicating there is no straightforward inference to be made from judgments about existential cases to universal cases. However, in contrast to Latham and Tierney (2021), we also find evidence that manipulation type is important too, though not in the way that opponents of manipulation arguments have typically thought. We find that manipulation type influences the strength of the existential/universal effect in people’s free will judgments, but not in their moral responsibility judgements. This indicates that judgments about free will and moral responsibility can come apart under certain conditions, which puts pressure on an underlying assumption in the literature that judgments about free will and moral responsibility stand and fall together.

1. **Manipulation Type vs. Manipulation Scope**

There are several manipulation arguments against compatibilism in the literature. Take, for example, Peter van Inwagen’s (1983) objection to Antony Flew’s (1955) Paradigm Case Argument for compatibilism. Van Inwagen presents the following case:

(M) When any human being is born, the Martians implant in his brain a tiny device—one that is undetectable by any observational technique we have at our disposal, though it is not *in principle* undetectable—which contains a “program” for that person’s entire life: whenever that person must make a decision, the device *causes* him to decide one way or the other according to the requirements of a table of instructions that were incorporated into the structure of the device before the person was conceived. (1983: 109)

Next, van Inwagen argues:

If we should discover that some particular person – Himmler, say – acted as he did because of a Martian device, implanted in his brain at the moment of his birth, had caused all his decisions, then we should hardly want to say that Himmler had free will, that he could have helped what he did, that he had any choice about the way he acted, or that he ever could have done otherwise. (1983: 110)

Van Inwagen first argues that agents, if they were to discover that Himmler was manipulated by Martians, would judge him to be unfree. This, for van Inwagen, is obvious. In fact, he takes the claim that free will is compatible with (M) to be “absurd” (1983: 112). From here, van Inwagen goes on to argue that the compatibilist can only save their view if they are able to identify a relevant difference between (M) and determinism, such that Himmler is not free while those who occupy deterministic universes still can be. But van Inwagen is not confident that compatibilists could succeed. Importantly, he does not deny that there are many differences between (M) and determinism: For example, (M) entails that an agent’s decisions to act are chosen by non-human entities and caused by devices that are, quite literally, alien to them, while the truth of determinism entails no such thing (1983: 111). However, according to van Inwagen, while these differences are important, because “they are what makes it *evident* that (M) is incompatible with free will…” (1983: 111), they are not relevant—they give us no reason to think that (M) and determinism bear different relationships to free will. So, van Inwagen concludes, determinism, like (M), is incompatible with free will, and Flew’s Paradigm Case Argument for compatibilism fails.

 Other theorists who develop manipulation arguments against compatibilism make an even stronger claim about the relationship between manipulation and determinism. According to these theorists, the reason we find agents in manipulation cases to be unfree is *because* their decisions to act are determineddue to the manipulation. Call this the Determinism Hypothesis.

**Determinism Hypothesis**: if an agent *A* is causally determined by factors that are beyond their control to decide to perform action *X*, then we judge *A* to be unfree and not responsible for deciding to perform *X*.

The Determinism Hypothesis plays a central role in many contemporary manipulation arguments. In Pereboom’s four-case argument (Pereboom 2001, 2014), Pereboom first presents three cases in which an agent, Plum, is causally determined to decide to kill White by radio-like technology, neuroscientific programming, and community training. He then presents a fourth case in which Plum is causally determined to decide to kill White by the past facts and laws of nature (i.e., a case in which determinism is true). According to Pereboom, the Determinism Hypothesis is what correctly explains our judgment that Plum is not free and morally responsible in Cases 1–3. Thus, consistency demands that we should also judge that Plum is not free and responsible in Case 4. If that is right, then free will and responsibility are not compatible with determinism, and so compatibilism is false.

 Mele (2006) argues similarly using his zygote argument. Mele first presents a case in which a powerful goddess intentionally creates a zygote within a deterministic universe to ensure that they will develop into the agent Ernie that will perform a particular action in thirty years’ time. Mele then argues that intuitively Ernie is not free and responsible due to the way their zygote was produced, and that these conditions are relevantly like determinism. If the Determinism Hypothesis correctly explains our judgment that someone causally determined by a powerful goddess is not free and responsible, then consistency demands that we should also judge the same if they are causally determined by the past facts and laws of nature. And to judge that causally determined agents are un-free and not morally responsible is to reject the truth of compatibilism.

 While theorists like van Inwagen, Pereboom, and Mele take there to be no relevant differences between manipulation, as featured in their manipulation cases, and the truth of determinism, many compatibilists have disputed this. They argue that there is in fact a relevant difference between manipulation and determinism cases, typically pointing to the presence of a *manipulator.* According to these arguments, the reason we find agents unfree in manipulation cases is not because their actions are determined full stop, but rather because they are *determined by another agent*. Call this the Non-Intentional Manipulator Hypothesis.[[5]](#endnote-5)

**Non-Intentional Manipulator Hypothesis**: if an agent *A* is causally determined by other agents to decide to perform action *X*, then we judge *A* to be unfree and not responsible for deciding to perform *X.*

Some compatibilists defend a narrower manipulation hypothesis, arguing that the presence of a manipulator only undermines free will and moral responsibility if the manipulator *intends* that an agent makes a decision to perform a particular action and then intervenes to ensure that they do so (e.g., Deery & Nahmias 2017; Waller 2014). Call this the Intentional Manipulator Hypothesis.

**Intentional Manipulator Hypothesis**: if an agent *A* is causally determined by other agents to decide to perform action *X*, and these agents intend for A to decide to perform action *X*, then we judge *A* to be unfree and not responsible for deciding to perform *X.*

If either manipulator hypothesis is right, then the reason we find agents in manipulation cases unfree and not responsible has nothing to do with determinism, but rather with their actions being the result of either non-intentional or intentional manipulation. What matters, according to these compatibilists, is not that our actions are causally determined, but *what* does the causal determining. One can grant that manipulators undermine free will and moral responsibility without granting that determinism does so.

 Recently, Latham and Tierney (2021, 2022) identified another potentially relevant difference between manipulation and determinism that could account for the intuition that agents in manipulation cases are not free and responsible. Notice that in van Inwagen’s, Pereboom’s, and Mele’s manipulation cases, a single agent—Himmler, Plum, and Ernie, respectively—is the target of *existential manipulation*: he is the onlyagent in his universe that is manipulated. Latham and Tierney predicted that existential manipulation drives the intuition that agents in manipulation cases are not responsible for their decisions to act. Call this the Existential Hypothesis.

**Existential hypothesis:** if an agent *A* is causally determined to decide to perform action *X* by a manipulation which only affects *A*, then we judge *A* to be unfree and not responsible for deciding to perform *X.*

If this hypothesis is right, then there may well be another relevant difference between manipulation and determinism, contrary to what the defenders of manipulation arguments claim. That’s because determinism, if true, does not affect only one agent. It is a universal thesis—its scope includes every agent, and everything, within a universe. Thus, the existential hypothesis does not extend to cases in which determinism is true.

 In the literature on manipulation arguments, the existential hypothesis has attracted far less attention than the determinism and manipulator hypotheses. This is puzzling, since the assumption that manipulation scope is irrelevant to our judgments of free will and moral responsibility is essential to many manipulation arguments. For these arguments to work, one must assume that existential manipulation is relevantly similar to universal determinism. Call this the Universal Defeater Principle.

**The Universal Defeater Principle**: if we judge that some factor *F* undermines free will and moral responsibility, then the discovery that *F* is true of all agents within a universe will lead to the judgment that no one is free and responsible.

Though crucial to the success of manipulation arguments, this principle is rarely articulated (let alone explicitly defended), though van Inwagen provides a notable exception. After arguing that we would not find Himmler free if we were to discover that he was implanted with the Martians’ device, van Inwagen claims: “And I don’t see why matters should be different if we discovered that *everyone* was “directed” by a Martian device: then we should have to make judgements about everyone” (1983: 110). Given the important role manipulation scope plays in manipulation arguments, we should investigate whether matters are in fact different in existential versus universal contexts.

 Happily, recent work in experimental philosophy directly bears on this question. Latham and Tierney (2021) tested people’s free will and moral responsibility judgments about existential and universal cases of non-intentional manipulation and determinism. They found, as predicted by the existential hypothesis, that people’s free will and responsibility judgments were significantly higher in response to universal cases than to existential cases. In contrast, they found *no* evidence that manipulation type had any effect on people’s judgments. If the determinism hypothesis were correct, then people should have judged that agents were equally unfree and responsible in the existential and universal manipulation cases, and equally unfree and responsible in the existential and universal determinism cases. This is because there was no difference with respect to determinism within these pairs of cases. Alternatively, if the non-intentional manipulator hypothesis were correct, then people should have judged that the agent was free and responsible in the determinism cases, but *not* the manipulation cases. This is because only the agent’s decision in the manipulation cases was causally determined by other agents. However, no such evidence was found. Taken together, Latham and Tierney’s (2021) results suggest that people’s free will and moral responsibility judgements are sensitive to manipulation scope, which puts pressure on the Universal Defeater Principle, but *not* manipulation type, which challenges the determinism and non-intentional manipulator hypotheses.

 While suggestive, Latham and Tierney’s (2021) experiment is limited in various respects. First, they only tested cases featuring one kind of manipulation—cultural conditioning—but not cases that featured less familiar forms of manipulation, like alien intervention. Second, Latham and Tierney did not test cases of manipulation involving intentional outcomes, so their study does not directly bear on the intentional manipulator hypothesis. Theorists that defend the intentional manipulator hypothesis (e.g., Deery & Nahmias 2017; Waller 2014) are only committed to thinking that free will and moral responsibility are undermined when agents are causally determined by manipulators who intend for them to decide to act in a particular way. It is open to these theorists to think that agents are both free and responsible in determinism cases and non-intentional manipulation cases. Still, Latham and Tierney’s results do provide evidence against the intentional manipulator hypothesis being a *complete* explanation of people’s judgments. After all, the intentional manipulator hypothesis makes no predictions about why we should observe the existential/universal distinction in these cases. Finally, another limitation of Latham and Tierney’s 2021 study is their measurement of moral responsibility judgments, which they discuss in the conclusion of their paper (2021: 9). Latham and Tierney simply asked participants to judge the degree to which Katie is morally responsible for her decision to kill Plum. But there are a variety of ways to understand what it means to be morally responsible, only some of which the incompatibilist takes to be at odds with the truth of determinism. Pereboom (2014), for example, takes *basic desert* moral responsibility—the extent to which agents truly deserve praise and blame for their actions—to be incompatible with the truth of determinism, but argues that forward-looking moral responsibility is not similarly threatened. Latham and Tierney’s (2021) measure of moral responsibility does not distinguish these multiple senses, making it difficult to draw conclusions.

 Thus, to fully assess the competing hypotheses, one must measure people’s judgments about free will, moral responsibility, anddeserved blame in response to existential and universal cases featuring determinism, non-intentional,and intentional manipulation. We ran an experiment to do exactly this, which we detail in the next section.

1. **The Experiment**

600 people were recruited online using Amazon Mechanical Turk. 187 participants were excluded for failing to respond to all the questions or answer the comprehension checks correctly. The final sample consisted of 413 participants (189 female, 3 trans/non-binary; aged 22-71; *M*=39.00, *SD*=10.89). Participants were randomly assigned to read one of six vignettes:[[6]](#endnote-6)

****

****

****

Following the vignette, participants were presented with three statements in randomized order: “Katie is morally responsible for killing Mr. Plum”, “Katie kills Mr. Plum of her own free will” and “Katie deserves blame for killing Mr. Plum.” Participants were asked to indicate their level of agreement for each statement on separate 7-point Likert scales which ran from 1 (*strongly* disagree) to 7 (*strongly* agree). The orientation of the Likert scales was randomized.

 On a separate screen, participants were then given four comprehension check questions in randomized order. Those that received a determinism vignette saw statements: (A) “The past facts and laws of nature completely determine Katie’s actions” and (B) “The past facts and laws of nature completely determine most people’s actions.” Those that received the intentional and non-intentional manipulation vignettes saw statements: (A) “An implanted alien device completely determines Katie’s actions” and (B) “Implanted alien devices completely determine most people’s actions.” All participants saw the statement (C) “Katie’s desire to kill Mr. Plum played a role in her deciding to kill Mr. Plum.” Participants were asked to indicate their level of agreement with each statement on a separate 7-point scale. Participants who received an existential vignette were excluded from the analyses if they failed to agree with Statement A (5, 6, 7) and disagree with Statement B (1, 2, 3). Participants who received a universal vignette were excluded from the analyses if they failed to agree with Statement A and Statement B. Participants were also excluded if they failed to agree with Statement C (5, 6, 7). Finally, participants saw the statement (D): “In this vignette, you were asked to imagine that Katie…” along with four response options: (a) “Did not want to kill Mr. Plum,” (b) “Was raised to make decisions that almost never benefit her,” (c) “Did not identify with the desire to kill Mr. Plum,” and (d) “Can regulate her behaviour by moral reasoning.” Participants who failed to choose option (d) were excluded from the analyses.

 We examined participants’ moral responsibility, free will, and blameworthiness judgments across conditions using a MANOVA. This analysis revealed a significant main effect of manipulation scope (existential, universal), Λ=.781, *F*(3, 405)=37.950, *p*<.001, and manipulation type (determinism, intentional manipulation, non-intentional manipulation), Λ=.843, *F*(6, 810)=12.074, *p*<.001, on people’s judgements. It also revealed a significant interaction effect between manipulation scope and manipulation type on people’s judgments, Λ=.910, *F*(6, 810)=6.549, *p*<.001. Next, we report the results of separate ANOVAs to show the effects of manipulation scope, manipulation type, and their interaction, on moral responsibility, free will, and blameworthiness judgments.

 Table 1 summarises the descriptive results for moral responsibility judgments. The %MR column is the proportion of participants who agree that Katie is morally responsible for killing Mr. Plum (5, 6, 7). The %~MR column is the proportion of participants who disagree that Katie is morally responsible for killing Mr. Plum (1, 2, 3). The ‘4’ column is the proportion of people who are indifferent to the statement. Separate one-sample t-tests were run for each condition to test whether mean levels of agreement significantly differed from 4 (indifference). Results of those tests (two right-hand columns) show that, *overall*, participants judged Katie morally responsible for killing Mr. Plum in each condition except for in the existential intentional and non-intentional manipulation conditions.[[7]](#endnote-7)



The results of the ANOVA revealed a significant main effect of manipulation scope, *F*(1, 407)=41.024, *p*<.001, manipulation type, *F*(2, 407)=11.977, *p*<.001, and significant two-way interaction between manipulation scope and manipulation type, *F*(2, 407)=7.707, *p*<.001. The significant main effect of manipulation scope was that people’s moral responsibility judgments in universal conditions (*M*=5.70, *SD*=1.52) were significantly higher than in existential conditions (*M*=4.73, *SD*=1.53). Pairwise comparisons using a Bonferroni correction for the main effect of manipulation type showed that people’s moral responsibility judgments in determinism conditions (*M*=5.71, *SD*=1.56) were significantly higher than in intentional manipulation conditions (*M*=4.80, *SD*=1.53; *p*<.001), and non-intentional manipulation conditions (*M*=5.14, *SD*=1.56; *p*=.008). There was no significant difference in people’s moral responsibility judgments between the intentional and non-intentional manipulation conditions (*p*=.204). Simple effects tests using a Bonferroni correction were carried out on the two-way interaction between manipulation scope and manipulation type.

 First, for existential conditions, people’s moral responsibility judgments in the determinism condition were significantly higher than in the intentional manipulation condition (*p*<.001) and non-intentional manipulation condition (*p*=.002). There was no significant difference in people’s moral responsibility judgments between the intentional manipulation and non-intentional manipulation conditions (*p*=.066). Second, for universal conditions, there was no significant difference in people’s moral responsibility judgments between the determinism, intentional manipulation, and non-intentional manipulation conditions (*p*>.999). Third, for determinism conditions, there was no significant difference in people’s moral responsibility judgments between the existential and universal conditions (*p*=.428). Finally, for intentional manipulation and non-intentional manipulation conditions, people’s moral responsibility judgments in the existential conditions were significantly lower than in the universal conditions (*p*<.001).

Table 2 summarises the descriptive results for blameworthiness judgments and follows the same organizational pattern as Table 1. One-sample t-test results show that, *overall*, participants judged Katie blameworthy for killing Mr. Plum in each condition *except* the existential intentional and non-intentional manipulation conditions.



The ANOVA revealed a significant main effect of manipulation scope, *F*(1, 407)=38.151, *p*<.001, manipulation type, *F*(2, 407)=21.663, *p*<.001¸and significant two-way interaction between manipulation scope and manipulation type, *F*(2, 407)=14.392, *p*<.001. The significant main effect of manipulation scope was that people’s blameworthiness judgments in universal conditions (*M*=5.68, *SD*=1.52) were significantly higher than in existential conditions (*M*=4.74, *SD*=1.52). Pairwise comparisons using a Bonferroni correction for the main effect of manipulation type showed that people’s blameworthiness judgments in determinism conditions (*M*=5.90, *SD*=1.54) were significantly higher than in intentional manipulation conditions (*M*=4.73, *SD*=1.52; *p*<.001) and non-intentional manipulation conditions (*M*=4.99, *SD*=1.54; *p*<.001). There was no significant difference in people’s blameworthiness judgments between the intentional manipulation and non-intentional manipulation conditions (*p*=.466). Simple effects tests using a Bonferroni correction were carried out on the two-way interaction between manipulation scope and manipulation type.

 First, for existential conditions, people’s blameworthiness judgments in the determinism condition were significantly higher than in the intentional manipulation condition (*p*<.001) and non-intentional manipulation condition (*p*<.001). There was no significant difference in people’s blameworthiness judgments between the intentional manipulation and non-intentional manipulation conditions (*p*=.183). Second, for universal conditions, there was no significant difference in people’s blameworthiness judgments between the determinism, intentional manipulation, and non-intentional manipulation conditions (*p*>.999). Third, for determinism conditions, there was no significant difference in people’s blameworthiness judgments between the existential and universal condition (*p*=.512). Finally, for intentional manipulation and non-intentional manipulation conditions, people’s blameworthiness judgments in the existential conditions were significantly lower than in the universal conditions (*p*<.001).

 Table 3 summarises the descriptive results for free will judgments and follows the same organizational pattern as Tables 1 and 2. One-sample t-test results show that, overall, participants think that Katie freely kills Plum in each condition *except* the existential intentional and non-intentional manipulation conditions. Overall, participants think that Katie does *not* freely kill Plum in the existential intentional manipulation condition.



The results of the ANOVA revealed a significant main effect of manipulation scope, *F*(1, 407)=112.418, *p*<.001, manipulation type, *F*(2, 407)=29.352, *p*<.001¸and significant two-way interaction between manipulation scope and manipulation type, *F*(2, 407)=10.617, *p*<.001. The significant main effect of manipulation scope was that people’s free will judgments in universal conditions (*M*=5.49, *SD*=1.62) were significantly higher than in existential conditions (*M*=3.78, *SD*=1.61). Pairwise comparisons using a Bonferroni correction for the main effect of manipulation type showed that people’s free will judgments in determinism conditions (*M*=5.35, *SD*=1.65) were significantly higher than in intentional manipulation conditions (*M*=3.85, *SD*=1.62; *p*<.001), and non-intentional manipulation conditions (*M*=4.69, *SD*=1.65; *p*=.003). People’s free will judgments in intentional manipulation conditions were significantly lower than in non-intentional manipulation conditions (*p*<.001). Simple effects tests using a Bonferroni correction were carried out on the two-way interaction between manipulation scope and manipulation type.

 First, for existential conditions, people’s free will judgments in the determinism condition were significantly higher than in the intentional manipulation condition (*p*<.001) and non-intentional manipulation condition (*p*=.024). People’s free will judgements in the intentional manipulation condition were significantly lower than in the non-intentional manipulation condition (*p*<.001). Second, for universal conditions, people’s free will judgments in the determinism condition were significantly higher in the intentional manipulation condition (*p*=.034), but not the non-intentional manipulation condition (*p*=.135). There was no significant difference in people’s free will judgments between the intentional manipulation and non-intentional manipulation conditions (*p*>.999). Finally, for determinism, intentional manipulation, and non-intentional manipulation conditions, people’s free will judgments in existential conditions were significantly lower than in universal conditions (*p*<.001).

1. **Results**

In this section, we summarize our major findings with respect to each of the tested hypotheses.

*Determinism Hypothesis*

According to the determinism hypothesis, the reason we find agents in manipulation cases to be lacking free will and moral responsibility is because these agents’ decisions are determined by factors that are beyond their control. Thus, the determinism hypothesis predicts that participants in all of our manipulation conditions will find Katie to be unfree and not responsible or blameworthy, since Katie’s choice to kill Mr. Plum was determined by factors beyond her control in each case.[[8]](#endnote-8) Participants, though, tended to judge that Katie was free and responsible in each case. There was also a significant difference in people’s judgments in response to existential and manipulation conditions, which the determinism hypothesis fails to predict.

 Interestingly, however, we failed to find any evidence of an existential/universal effect on people’s moral responsibility judgments in the determinism conditions. Importantly, though, the lack of an effect is not because people judge that Katie is equally *not* responsible for her choice to kill Mr. Plum in the existential and universal determinism cases. Rather, it is because people judge that she is *equally* responsible in both cases. The fact that participants’ judgments of moral responsibility were not sensitive to either existential or universal determinism suggests that determinism is *not* a factor in the formation of these judgments. In short, people fail to treat determinism as a threat to moral responsibility, despite what the determinism hypothesis predicts and what many incompatibilists have assumed.

 The fact that we failed to find the existential/universal effect with respect to participants’ judgments of moral responsibility in response to the determinism cases is noteworthy for several reasons. First, this result differs from Latham and Tierney’s (2021) findings and indicates that the scope of the existential/universal effect may be narrower than they initially proposed. Perhaps some manipulations have *no* impact on our judgments of responsibility, even in their existential form. This result also poses a serious challenge to defenders of manipulation arguments, for it indicates that there is no straightforward route of argumentation from existential manipulation cases to determinism cases, even in their existential form. Not only did participants find Katie to be morally responsible in the existential determinism case, but they also found her to be significantly more responsible than in both kinds of existential manipulation cases. This suggests that there is a relevant difference between existential manipulation and existential determinism, such that even if we observed that existential manipulation cases undermine moral responsibility, this would not provide any warrant for thinking the same of existential (or universal) determinism.[[9]](#endnote-9)

*Manipulator Hypotheses*

The fact that determinism did not have a measurable effect on participants’ judgments of free will and moral responsibility is consistent with both the intentional and non-intentional manipulator hypotheses. However, we did not find further support for either of these hypotheses. According to the non-intentional manipulator hypothesis, if an agent is causally determined by other agents to decide to perform an action, then we judge that the agent is not free and not responsible. And according to the intentional manipulator hypothesis, we judge that the agent is not free and not responsible only if the manipulator also intended that the agent decide to perform the action. While we found evidence that people’s moral responsibility and free will judgments were sensitive to the presence of manipulators (intentional or not), the way this sensitivity presented itself does not support the non-intentional and intentional manipulator hypotheses. (We will reflect on the way in which participants’ judgments were sensitive to manipulation kind shortly). If the non-intentional or intentional manipulator hypotheses were correct, then we should have observed that participants judge Katie to be equally unfree and not responsible in existential and universal non-intentional manipulation cases and/or existential and universal intentional manipulation cases. But people judged, overall, that Katie was not free and responsible only in the existential non-intentional and intentional manipulation cases, and judged, overall, that Katie was free and responsible in universal versions of these cases.

 Neither the non-intentional nor intentional manipulator hypothesis predicts, or can account for, the existential/universal effect observed in these conditions. Instead, what this result shows is that there is no straightforward route of argumentation from existential cases to universal cases, even within the same manipulation type. This suggests that there is a relevant difference to people between existential manipulation and universal manipulation, such that even if we observed that existential manipulation cases undermine free will and moral responsibility, this would not provide warrant for thinking the same of universal manipulation cases.

*Existential Hypothesis*

In contrast with the determinism, non-intentional manipulator, and intentional manipulator hypotheses, the existential hypothesis can account for most of the findings in this study. According to the existential hypothesis, if an agent is causally determined to perform some action by a manipulation that affects only that agent, then we will find them unfree and not responsible for that action. Overall, this is precisely what we found—participants judged Katie to be significantly less free and responsible in existential cases than in universal cases. The only exception to this was people’s moral responsibility judgments about determinism cases, discussed above.

 We also replicated Latham and Tierney’s finding that the existential/universal effect is larger for free will judgements than for moral responsibility judgments (2021: fn. 11, 14). We are now able to see why: the existential/universal effect for free will judgments is associated with manipulation type, whereas the existential/universal effect for moral responsibility judgments is not. More specifically, the existential/universal effect for free will judgments is largest in intentional manipulation cases, and smallest in determinism cases, with non-intentional manipulation in between. In contrast, there was no existential/universal effect for moral responsibility judgments about determinism cases, and there was no difference in the size of the existential/universal effect for intentional and non-intentional manipulation cases. Thus, our results largely replicate Latham and Tierney’s (2021) findings and build on them in several interesting ways.

 Our study addresses potential limitations of Latham and Tierney’s (2021) study as well. For example, unlike Latham and Tierney (2021), we measured participants’ judgments about both moral responsibility and blameworthiness. If people are sensitive to the distinction between basic desert responsibility and other forms of responsibility, and if these judgments are impacted differently by the factors examined in this study, then we would expect to observe different patterns in people’s moral responsibility and blameworthiness judgments. However, we found no evidence of any differences between people’s moral responsibility and blameworthiness judgments. This suggests that Latham and Tierney’s (2021) original findings are not due to basic desert responsibility judgements and other forms of responsibility judgements being confounded. Of course, it is possible that people might still not be operating with the concepts of ‘moral responsibility’ and ‘blameworthiness’ that incompatibilists take to be threatened by the truth of determinism. To rule out this possibility, it could be useful to tutor participants on the precise notions of responsibility and blameworthiness each term is meant to pick out, and only after having done this should we expect to accurately measure people’s judgements. While this is worth exploring in future research, it seems likely that people simply think that blameworthiness, and thus basic desert responsibility, is compatible with the truth of determinism.

1. **Discussion**

The results of this study suggest that manipulation scope impacts our judgments of free will and moral responsibility. But *should* our judgments be affected in this way? Manipulation scope has no direct bearing on agents, or their abilities, and is rather simply a feature of agents’ circumstances and how they compare to other agents in their world. Can theories of free will and moral responsibility account for these purely relational and circumstantial facts, and would those who defend these views be motivated to do so?

 Theorists who take free will and moral responsibility to be grounded in deep metaphysical facts about agents alone will likely reject the relevance of manipulation scope to these phenomena. But many accounts of free will and moral responsibility, particularly those defended by compatibilists, already make room for facts about agents’ circumstances and their relationships with others.[[10]](#endnote-10) The theorists who defend these views will be better placed, and perhaps more motivated, to account for the role manipulation scope plays in determining the degree to which agents are free and responsible. In fact, Latham and Tierney (2022) recently argued that Strawsonian approaches to moral responsibility, which take facts about our responsibility practices to be deeply connected to the nature of responsibility itself, can account for the existential/universal effect, since our practices are impacted in very different ways by existential and universal phenomena.

 Because studies like ours shed light onto our responsibility practices, and because Strawsonian views take the nature of moral responsibility to be informed by these practices, one could expect Strawsonians to be concerned with accommodating our results. But recall that our study indicates that manipulation scope affects judgments about free will in different ways than it affects judgments about moral responsibility. Could a Strawsonian approach accommodate this pattern of judgments? We think it could but doing so might require revision. Many Strawsonians study our praising and blaming practices to understand the nature of moral responsibility and attend less directly to questions about free will. But if it turns out that certain factors affect our attributions of free will but not our attributions of moral responsibility (or vice versa), then Strawsonians would do well to attend to how our responsibility practices come apart from our free will practices. Only by fully examining both sets of practices will the Strawsonian be able to offer complete practice-first accounts of these phenomena. And a full exploration of the differences between our moral responsibility and free will practices might require Strawsonians, as well as other compatibilists, to revise other elements of their theories. For example, some Strawsonians, in addition to compatibilists of other stripes, take free will to be necessary for moral responsibility (e.g., McKenna 2012: 13). But if it turns out that certain factors, like existential intentional manipulation, undermine our judgments of free will, but have no significant effect on our assessments of moral responsibility, then practice-first accounts might need to significantly revise how they understand free will to relate to moral responsibility. We take the exploration of how our moral responsibility and free will practices interact, and the potential revisions that such an investigation could lead to, to be fruitful directions for future research.

Of course, even if theorists *can* accommodate relational and circumstantial factors like manipulation scope into their views of free will and moral responsibility, this doesn’t mean that they *should*. Our study indicates that universal manipulation does not threaten judgments of free will and moral responsibility, but many theorists might not be willing to accept these results. They might balk at the thought that there could be free will and moral responsibility in a world where everyone is causally determined by implanted alien control devices, for example. However, there is an important distinction to be made here, which is not often made in the literature, between something being necessary, and something being necessary because it’s actual. One surprising theorist who makes this distinction is van Inwagen. At the end of *An Essay on Free Will*, he writes: “it is conceivable that science will one day present us with compelling reasons for believing in determinism. Then, and only then, I think we should become compatibilists (1983: 223).” According to van Inwagen, conditional on determinism being true, certain features of agents and their actions can ground free will and moral responsibility, whereas if determinism is false, these features would be irrelevant to free will and moral responsibility. The idea that our concept of free will takes a conditional form with respect to determinism, while underexplored, has been articulated in the literature (Björnsson 2014, Latham 2019; Roskies & Nichols 2008). On this approach, if the actual world is indeterministic, and agents have libertarian powers, then libertarian powers are what free will and moral responsibility is and necessarily so. However, if the actual world is deterministic, and agents have access to only compatibilist powers, then compatibilist powers are sufficient for free will and moral responsibility.

 If theorists, including van Inwagen, are willing to be compatibilists when presented with the truth of universal determinism, then why should we not become compatibilists with respect to the truth of other purported universal threats to our free will and responsibility as well? Of course, this is not to say that anything will go, but given the fact that people overwhelmingly agree that agents are free and responsible even if everyone’s decisions are causally determined by implanted alien devices, the kinds of powers that might count as acceptable, if actual, might be very minimal. Even in a world that by our lights would be hostile to freedom and responsibility, if such a world was discovered to be actual, we might be disposed to judge that we can still be free and responsible.

1. **Conclusion**

In this paper, we built on Latham and Tierney’s (2021) study, which first identified the impact of the existential/universal effect on judgments of free will and moral responsibility. We replicated their findings in a wider and more diverse range of cases while also revealing a more complex picture of our responsibility and free will judgments than has previously been provided. Notably, the lack of evidence of an existential/universal effect in people’s moral responsibility judgments about determinism cases suggests that people do *not* think that determinism is a threat to responsibility and think that it is relevantly *unlike* manipulation. Furthermore, the presence of the existential/universal effect in both intentional and non-intentional manipulation cases suggests that while people think that existential manipulation undermines free will and moral responsibility, this does *not* provide any warrant for thinking the same of universal cases. Finally, the fact that we found the size of the existential/universal effect to be tied to manipulation type in the case of free will judgments, but not in the case of moral responsibility judgments, suggests that these judgments could come apart in some contexts.

 Our results also raise interesting questions that should be pursued in future research. As discussed above, given the differences found between judgments of free will and moral responsibility, philosophers would do well to both conceptually and empirically attend more carefully to how free will and moral responsibility relate to one another. It will also be important to explore the bounds of the existential/universal effect on our judgments of free will and moral responsibility. In the current study, we examined cases that uniformly featured morally bad actions. It would be worthwhile to examine whether the existential/universal effect also arises in positively valenced contexts and impacts judgments of praiseworthiness. Some theorists argue that there are important asymmetries between praise and blame (Nelkin 2011; Wolf 1990) and interesting differences between judgments of praiseworthiness and blameworthiness have been empirically found (Knobe 2003, 2006). Thus, it’s an open question as to whether manipulation scope impacts our judgments of praiseworthiness and moral responsibility for positively valenced actions in the same way it impacts judgments of blameworthiness. We hope future research will shed light on these important questions.

**References**

Björnsson, G. (2014). Incompatibilism and “Bypassed” Agency. In A. R. Mele (ed.), *Surrounding Free Will*. OUP.

Deery, O. and Nahmias, E. (2017). Defeating Manipulation Arguments: Interventionist Causation and Compatibilist Sourcehood. *Philosophical Studies* 174(5): 1255–1276.

Flew, A. (1955). Divine Omnipotence and Human Freedom. In Flew, A. and McIntyre, A. (ed.) *New Essays* *in Philosophical Theology*. SCM Press.

Kane, R. (1996). *The Significance of Free Will*. OUP.

Knobe, J. (2003). Intentional Action and Side-Effects in Ordinary Language. *Analysis* 63:

 160–163.

Knobe, J. (2006). The Concept of Intentional Action: A Case Study in the Uses of Folk

 Psychology. *Philosophical Studies* 130: 203–231.

Latham, A.J. (2019). The Conceptual Impossibility of Free Will Error Theory. *European Journal of Analytic Philosophy* 15(2): 99–120.

Latham, A. J., & Tierney, H. (2021). The Four-Case Argument and the Existential/Universal Effect. *Erkenntnis* 1–11.

Latham, A. J., & Tierney, H. (2022). Defusing Existential and Universal Threats to Compatibilism: A Strawsonian Dilemma for Manipulation Arguments. *Journal of Philosophy* *119*(3): 144–161.

Markosian, N. (2012). Agent Causation as the Solution to All the Compatibilist’s Problems. *Philosophical Studies* *157*(3): 383–398.

McKenna, M. (2012). *Conversation and Responsibility*. OUP.

Mele, A. R. (2006). *Free Will and Luck.* OUP.

Nelkin, D. (2011). *Making Sense of Freedom and Responsibility*. OUP.

Nichols, S., & Knobe, J. (2007). Moral Responsibility and Determinism: The Cognitive Science of Folk Intuition. *Noûs* 41: 663–682.

Pereboom, D. (2001). *Living Without Free Will*. CUP.

Pereboom, D. (2014). *Free will, Agency, and Meaning in Life*. OUP.

Roskies, A. and Nichols, S. (2008). Bringing Moral Responsibility Down to Earth. *Journal*

 *of Philosophy* 105: 371–388.

Schlosser, M. (2015). Manipulation and the Zygote Argument: Another Reply. *Journal of Ethics* 19(1): 73–84.

Van Inwagen, P. (1983). *An Essay on Free Will*. OUP.

Waller, R. R. (2014). The Threat of Effective Intentions to Moral Responsibility in the Zygote Argument. *Philosophia* 42: 209–222.

Wolf, S. (1990). *Freedom within Reason.* OUP.

1. See van Inwagen 1983; Kane 1996; Mele 2006; Pereboom 2001, 2014. [↑](#endnote-ref-1)
2. Here we are *only* using ‘manipulation’ as a category label. It’s up for grabs whether determinism is a form of, or is relevantly similar to, manipulation. [↑](#endnote-ref-2)
3. Again, we are using ‘manipulation type’ as a category label. It’s up for grabs whether any of the phenomena falling under ‘manipulation type’ are in fact manipulations. [↑](#endnote-ref-3)
4. See Deery & Nahimas 2018; Schlosser 2014; Waller 2014. [↑](#endnote-ref-4)
5. For ease of presentation, we refer to manipulation without an intentional outcome as ‘non-intentional manipulation’ and manipulation with an intentional outcome as ‘intentional manipulation.’ Likewise, we refer to manipulators who do not intervene in order to bring about a particular outcome as ‘non-intentional manipulators’ and manipulators who intervene in order to bring about a particular outcome as ‘intentional manipulators.’ [↑](#endnote-ref-5)
6. “Has to” is one expression that has been used in experimental philosophy to characterize the modal commitments of determinism for lay audiences (e.g., Nichols and Knobe 2007). However, one might worry about the use of this expression. “Has to” is ambiguous and is commonly used to describe being bound by norms and expectations (e.g., “has to” tell the truth, “has to” make a deadline, and so on). This ambiguity makes “has to” particularly friendly to incompatibilist intuitions. We agree that “has to” might have made our vignettes friendly to incompatibilist intuitions, but this renders our results even more interesting, since most participants responded in a *compatibilist*-friendly manner. Nevertheless, this result could also be due to something particular about our vignettes. Thus, in future work it will be important to rerun the study using different descriptions of determinism and manipulation. Thanks to Michael McKenna for discussion on this point. [↑](#endnote-ref-6)
7. Using a Bonferroni adjustment for multiple comparisons. [↑](#endnote-ref-7)
8. Interestingly, there were no differences in results between people’s moral responsibility and blameworthiness judgments. So, for ease of presentation, when we talk about people’s moral responsibility judgments, we are also talking about people’s blameworthiness judgments, unless otherwise specified. At the end of this section, we reflect on the lack of evidence of a difference between these two judgment types. [↑](#endnote-ref-8)
9. This result could also require Latham and Tierney to modify their view, for in other work (2022) they suggest that existential determinism could mitigate judgments of moral responsibility. They write: “If we come to know that a single individual in our universe is subject to determinism while every other agent is not, this may indeed affect our reactions to this individual and the relationships we cultivate with them” (2022: 156). They (2022) then argue that even if existential determinism mitigates moral responsibility, this does not require us to accept that incompatibilism is true, for incompatibilism is a thesis about the relationship between moral responsibility and *universal* determinism, and one cannot draw conclusions about a universal phenomenon from observations of existential phenomena—the two operate differently. Our results suggest that Latham and Tierney need not grant that existential determinism poses a threat to moral responsibility judgments. Rather, they could simply argue that determinism, in both its existential and universal form, does not impact moral responsibility judgments. In this way, the two phenomena *are* relevantly similar, though notably, in the very opposite respect that the determinism hypothesis predicts. [↑](#endnote-ref-9)
10. This isn’t to say that all compatibilists take this route. See, for example, Markosian (2012). [↑](#endnote-ref-10)