PHILOSOPHICAL PSYCHOLOGY
2020, VOL. 33, NO. 7, 968–996
https://doi.org/10.1080/09515089.2020.1817884

Routledge
Taylor & Francis Group

ARTICLE

Folk intuitions and the conditional ability to do otherwise

Thomas Nadelhoffer[a], Siyuan Yin[b] and Rose Graves[c]

[a]Philosophy, College of Charleston, Charleston, USA; [b]Kenan Institute for Ethics, Duke University, Durham, USA; [c]Statistical Science, Duke University, Durham, USA

ABSTRACT

In a series of preregistered studies, we explore (a) the difference between people's intuitions about indeterministic scenarios and their intuitions about deterministic scenarios; (b) the difference between people's intuitions about indeterministic scenarios and their intuitions about neuro-deterministic scenarios (i.e., scenarios where the determinism is described at the neurological level); (c) the difference between people's intuitions about neutral scenarios (e.g., walking a dog in the park) and their intuitions about negatively valenced scenarios (e.g., murdering a stranger); and (d) the difference between people's intuitions about free will and responsibility in response to first-person scenarios and third-person scenarios. We predicted that once we focused participants' attention on the two different abilities to do otherwise, available to agents in indeterministic and deterministic scenarios, their intuitions would support natural incompatibilism – the view that laypersons tend to judge that free will and moral responsibility are incompatible with determinism. This prediction is borne out by our findings.

ARTICLE HISTORY
Received 10 April 2019
Accepted 7 January 2020

KEYWORDS
Free Will; incompatibilism; Compatibilism; determinism; experimental Philosophy

1. Introduction

In an indeterministic universe, agents can have the unconditional ability to do otherwise – that is, they could have done other than they did, even if everything leading up to their decision remained the same. Libertarians suggest that this unconditional ability to do otherwise in the actual sequence of events is necessary for free will and responsibility – an ability that all parties to the free will debate agree is incompatible with a deterministic universe. After all, if determinism is true, given the very distant past and the physical laws, one and only one thing is possible at any given moment. In a deterministic universe, agents merely have the conditional ability to do otherwise – that is, agents could have acted differently only insofar as the past and the laws had been different than they actually are. Compatibilists suggest that this conditional ability to do otherwise (along with other

CONTACT Thomas Nadelhoffer tnadelhoffer@gmail.com Philosophy, College of Charleston, Charleston, USA

Supplemental data for this article can be accessed here.

cognitive and volitional capacities) can ground free will and moral responsibility; incompatibilists disagree.[1]

This feature of the free will debate – that is, the difference between the unconditional and the conditional ability to do otherwise – has been underexplored in the empirical literature on free will beliefs.[2] Our goal is to address this lacuna. So, in a series of preregistered, between-subject vignette-based studies, we explored (a) the difference between people's intuitions about indeterministic scenarios and their intuitions about deterministic scenarios; (b) the difference between people's intuitions about indeterministic scenarios and their intuitions about neuro-deterministic scenarios (i.e., scenarios where the determinism is described at the neurological level); (c) the difference between people's intuitions about neutral scenarios (such as when an agent walks his dog in the park) and their intuitions about negatively valenced scenarios (such as when an agent murders a stranger); and (d) the difference between people's intuitions about free will and responsibility in response to first-person scenarios and third-person scenarios. We predicted that once we focused participants' attention on the two different abilities to do otherwise, available to agents in indeterministic and deterministic scenarios, their intuitions would largely support natural incompatibilism – the view that most laypersons tend to judge that free will and moral responsibility are incompatible with determinism. As we will see, this prediction is borne out by our findings.

2. Natural compatibilism versus natural incompatibilism[3]

It has been commonplace for both parties to the free will debate to assume the mantle of common sense for their respective theories of free will. While folk intuitions need not fully constrain our theorizing, they can and should sometimes serve as the starting point of our philosophical investigations. Consequently, there is an ongoing debate about whether natural compatibilism or natural incompatibilism is the commonsense view – which is ultimately an empirical issue to be decided by systematic investigation. However, until experimental philosophers began exploring this issue, there was a dearth of data on what people actually think about the relationship between free will, responsibility, and determinism.

In some of the earliest work on this front, Nahmias et al. (2005, 2007)) set out to shed light on the relevant folk intuitions. Using three different descriptions of determinism, they found that a significant majority of participants (typically 65–85%) judged that agents in deterministic scenarios are free and morally responsible. These findings were constant across cases involving neutral actions (e.g., going jogging), positively valenced actions (e.g., saving a child from a burning building), and negatively valenced actions (e.g., robbing a bank), providing some intriguing preliminary data

that people are more compatibilist than incompatibilists have traditionally assumed. However, as is often the case in the free will debate, things quickly turned out to be more complicated.

For instance, Nichols and Knobe (2007) ran some follow-up studies to explore the psychological mechanisms that generate intuitions about moral responsibility. Participants were randomly assigned to either an "abstract" condition that describes a deterministic universe ("A") and an indeterministic universe ("B"), or a "concrete" condition that describes these universes but also describes a person in Universe A, Bill, who murders his wife and family in order to be with his secretary. Participants were then first asked which one of these two universes was more like their own. Nearly all participants (90%) answered "Universe B."[4] Then, participants in the abstract condition were asked whether it was possible for a person in Universe A to be "fully responsible for their actions." Participants in the concrete condition were asked instead, "is Bill fully morally responsible for killing his wife and children?" Whereas 72% of subjects gave the compatibilist response that Bill is fully morally responsible in the concrete condition, in the abstract condition, 84% gave the incompatibilist response that it is not possible for people in Universe A to be fully morally responsible.

On the surface, these findings appear to put pressure on the claim that people's intuitions are robustly compatibilist. Instead, whether people are inclined to give compatibilist answers may depend less on the presence (or absence) of determinism and more on the moral features of the vignettes. Whereas people tend to display compatibilist leanings when asked to make judgments concerning the responsibility of specific agents, when they are asked instead to think about responsibility in the abstract, their intuitions trend toward incompatibilism. There is an ongoing debate about how best to explain these competing findings. Nichols and Knobe take their results to show that while people have an incompatibilist theory of free will, when they consider a concrete situation, they make a performance error driven by the affectively charged nature of the scenario.

One of our primary goals is to contribute to this debate by constructing concrete deterministic scenarios that involve neutral actions to see whether, by focusing people's attention on the merely conditional ability to do otherwise in these scenarios, they would trend toward incompatibilism. We will say more about our experimental design in the next section. For now, we want to review the scant empirical literature on folk intuitions about the difference between the conditional and the unconditional ability to do otherwise – since that is directly relevant to the present task at hand. The first attempt to get at the salient intuitions was made by Nahmias et al. (2004). They gave participants the following survey:

> Imagine you've made a tough decision between two alternatives. You've chosen one of them and you think to yourself, "I could have chosen otherwise" (it may help if you can remember a particular example of such a decision you've recently made). Which of these statements best describes what you have in mind when you think, "I could have chosen otherwise"?
>
> (A) "I could have chosen to do otherwise even if everything at the moment of choice had been exactly the same."
> (B) "I could have chosen to do otherwise only if something had been different (for instance, different considerations had come to mind as I deliberated or I had experienced different desires at the time)."
> (C) Neither of the above describes what I mean.

The most popular choice was B (62%), followed by choices A (35%) and C (3%). In line with the rest of their work, Nahmias and colleagues took this to provide further support for natural compatibilism. After all, people don't appear to view choice and action through the lens of the unconditional ability to do otherwise – contrary to what incompatibilists have assumed. However, more recent work suggests that when care is taken to make sure that people understand the implications of determinism – which includes foreclosing on the unconditional ability to do otherwise – their intuitions about choice (and the experience of choice) are mostly incompatibilist (Deery et al., 2013). Over a series of three novel studies, Deery et al. (2013) consistently found that most people experience themselves as having the unconditional ability to do otherwise and that they take this ability to be inconsistent with determinism.[5]

In some follow-up work in developing and validating the Free Will Intuitions Scale (FWIS), Deery et al. (2015) further explored the different ways people might interpret the ability to do otherwise. They start their investigation by pointing out that the case method traditionally used by experimental philosophers has its limitations. On their view, this so-called "conflict method" – which forces people to rectify potentially competing intuitions by making forced choices – should be supplemented by an approach that relies on scales designed to measure people's background beliefs about free will and related constructs. While the conflict method reveals how people apply their underlying beliefs when forced to pass judgment on particular scenarios, the scale-based approach sheds light on the more stable underlying beliefs and attitudes themselves – beliefs which may be in tension in various ways with people's responses to scenarios. According to Deery et al. (2015), the scale-based method gets at people's "basic intuitions," whereas the conflict method gets at people's "decision" or "output" intuitions (p. 778). Because the conflict method only gets at the latter intuitions, it purportedly cannot shed light on the former intuitions – which is one of the goals of Deery and colleagues' adjunctive scale-based approach.[6]

In an attempt to access people's basic intuitions about free and moral responsibility, Deery and colleagues developed and validated the FWIS, which has eight subscales. The most important two subscales for present purposes are (a) ability-to-do-otherwise incompatibilism (e.g., "I could have decided to buy a different detergent than I actually bought, but I would have decided to do so even if none of my desires or thoughts at the time had been different"); and (b) ability-to-do-otherwise compatibilism (e.g., "I might have taken the job in Chicago instead of the job in Atlanta, but I would have done so only if my thoughts or desires had been different as I made the decision") (Deery et al., 2015, p. 796). Each of these subscales is designed to get at whether people's basic intuitions align with the conditional or unconditional ability to do otherwise. Their findings suggest that while people do distinguish between these two abilities, they tend to favor the former. However, while Deery and colleagues suggest that this preference is "strong," we don't think the data support such an assertion. Mean agreement with the conditional ability items was M= 5.06 (SD = 0.83), while mean agreement with the unconditional ability items was M = 4.27 (SD = 1.06). While this is surely a significant difference, we wouldn't characterize it as a contrastively strong preference. People's responses to the former are just above the midpoint and their responses to the latter are just above 'somewhat agree.' We think this is too mild of a difference to be adequately characterized as representing a strong preference. If someone said that they're undecided about x and someone else says they somewhat agree with x, it's odd to conclude that the latter has a strong preference for x relative to the former. On our reading, participants in their studies had a weak but significant preference for the conditional reading over the unconditional reading.

However, even if we set that disagreement aside, we have deeper worries with the way the items in the compatibilism subscale are worded. In each case, it sounds like the counterfactual possibilities were genuinely open at the time of the agent's decision – that is, it sounds to us as if the salient alternative beliefs, desires, and thoughts were live options. However, if determinism is true, given that the past and the laws are already set in stone, the agent's beliefs, desires, and thoughts could not have actually been different at the time. In this way, these items may unwittingly smuggle proximal indeterminism in through the back door – which is a potential serious weakness of this measure. For this subscale to be a measure of the compatibilist ability to do otherwise, Deery and colleagues would have needed to make it clear that at the moment these agents made their decisions, nothing could have actually been different, given the past and the laws (including the beliefs, desires, and thoughts of the agent). Otherwise, it is open for people to adopt an unconditional reading of

these purportedly conditional items – which we suspect partly explains their results.[7]

For present purposes, however, we want to bracket these concerns. If we instead take the findings by Deery et al. (2015) at face value, what should we make of the scale-based findings where the debate about natural compatibilism is concerned? In our eyes, not much. Why not? The reason is that the question that animates the debate about natural compatibilism is not what most people think about the ability to do otherwise in the indeterministic world they take themselves to inhabit, but what people think about this ability in a deterministic universe. Because participants in the Deery et al. (2015) studies are allowed to respond to the items of FWIS against the backdrop of their pre-theoretical commitment to indeterminism, their responses don't get at the issue that animates the debate about natural compatibilism – which instead has determinism as a backdrop. While the data yielded by the scale-based method are interesting and important in their own right, they don't shed much light on the issue at hand. Rather, the question we want to understand is what people think about the conditional and unconditional abilities to do otherwise when asked to assume that the universe is deterministic. While it is noteworthy that people who take themselves to be living in an indeterministic world will endorse, to varying degrees, both the conditional and unconditional abilities to do otherwise, that by itself does not speak to the debate about natural compatibilism. Even Deery and colleagues admit, the scale-based approach is a useful adjunct to the conflict method, but it cannot replace this method.

At the end of the day, if we want to get at whether people think that the conditional ability to do otherwise can undergird free and responsible agency in a deterministic universe, we have to present them with cases where these are the background conditions. This is precisely what we set out to do. Our goal is to build deterministic and indeterministic vignettes that emphasize the difference between the conditional and the unconditional ability to do otherwise. Our prediction was that by highlighting the fact that, in a deterministic scenario, agents merely have the conditional ability to do otherwise, people's intuitions would be broadly incompatibilist. As we will now discuss, this prediction is borne out by our findings.

3. Testing the conditional ability to do otherwise

3.1. Aims of present studies

We have several goals with these three studies. First, our primary goal is to explore people's intuitions about free will and moral responsibility in cases where the focus was on indeterministic scenarios involving the unconditional ability to do otherwise, or deterministic scenarios

involving the conditional ability to do otherwise. Here, we want to really emphasize what follows from these two types of abilities in terms of choice and action. To help accomplish this goal, we use a series of follow up statements for all three studies that enable us to clearly ascertain whether participants properly understood the deterministic implications of the scenario.[8] Second, we want to compare indeterministic scenarios with neuro-deterministic scenarios, that is, scenarios where the determinism is partly explained at the neurological level. Here, our motivation was the early work on neuro-prediction by Nahmias et al. (2007), which suggests that people find neuro-determinism more threatening to agency and responsibility than determinism per se.[9] By exploring people's intuitions about the conditional ability to do otherwise in neuro-deterministic scenarios, we will be able to broaden the scope of our investigation on this front. Third, we want to explore whether focusing on the conditional ability to do otherwise in deterministic universes would short-circuit or override the influence that affectively charged actions, like murder, have on people's intuitions about agency and responsibility. Here, we wanted to see whether once the full implications of determinism were made clear to participants, they would still find people free and responsible even in concrete situations involving immoral actions. Given that past research suggests that negatively valenced scenarios tend to prime compatibilist intuitions, if we can elicit incompatibilist intuitions by focusing on the conditional ability to do otherwise, this will strengthen our findings.

These three goals represent the main focus of our project since they enable us to explore both people's intuitions about free will and responsibility, on the one hand, as well as the conditional and unconditional abilities to do otherwise, on the other hand, across a variety of deterministic and indeterministic scenarios using both neutral and negatively valenced actions. We don't just want to use one type of scenario or one type of action; we want to get at a broader spectrum of intuitions. As a secondary interest, we also want to see whether people's intuitions across these cases would differ depending on whether the scenarios are worded in the first person or the third person. Here, we drew on the research on motivated social cognition and judgments about agency and responsibility (Clark et al., 2014; Lerner, 1980; Ross, 1977; Taylor & Brown, 1988). This second part of our project is not our core concern, but it was easy enough to add a self–other component to our experimental design, which is what we decided to do. While this self–other element doesn't directly speak to the divide between the conditional and unconditional ability to do otherwise per se, it does speak to the natural compatibilism debate more generally – since we predicted that people would be more incompatibilist when judging themselves and more compatibilist when judging others. Therefore, we thought it was

worth testing for this prediction while we were already administering our studies.

In light of these our goals, we designed a series of vignettes and follow-up statements, and we made the following predictions[10]:

- Attributions of free will and responsibility would be *lower* in the deterministic scenarios where agents merely have the conditional ability to do otherwise, than in the indeterministic scenarios where agents have the unconditional ability to do otherwise. This would provide evidence for natural incompatibilism.
- Attributions of free will and responsibility would be *lower* in the neuro-deterministic scenarios where agents merely have the conditional ability to do otherwise, than in the indeterministic scenarios where agents have the unconditional ability to do otherwise. This, too, would provide evidence for natural incompatibilism.
- Attributions of free will and responsibility would be *lower* in the negatively valenced neuro-deterministic scenario where the agent merely has the conditional ability to do otherwise, than in the negatively valenced indeterministic scenario where the agent has the unconditional ability to do otherwise. This, too, would provide evidence for natural incompatibilism.
- Attributions of moral responsibility would be *higher* than attributions of free will across conditions. This would provide evidence that free will is not simply whatever happens to undergird moral responsibility.
- Attributions of free will and responsibility would be *higher* in the first-person conditions than in the third-person conditions. This would provide evidence for the actor–observer bias.

Before we talk about the studies themselves, though, we want to say a few words about how we tried to operationalize determinism, which is the key issue we wanted to explore.

3.2. Settling on deterministic scenarios

A common way of defining determinism is to suggest that, in a deterministic universe, given the fixity of the past and the physical laws, at any moment, one and only one thing can happen. In this sense, the past and laws foreclose on actual, genuinely open possibilities in the stream of physical events. By "actual," we just mean keeping everything leading up to the moment of choice or action constant. Here, we are contrasting actual possibilities – the ones open to the agent at the moment of choice – from counterfactual possibilities – the ones that would have been open to the agent had things been slightly different in the distant past or with the laws.[11] While an

indeterministic universe is a "garden of forking paths," with genuinely open possibilities in the stream of actual events, in a deterministic universe, given the past and the laws, there is only one way things can unfold, like a train going down the tracks. Now, it's certainly true that in a deterministic universe, things could have been different – that is, the train could have broken down or headed down a different track – but only if something in the antecedent circumstances had been different (which they weren't). Compatibilists think this kind of possibility can partly undergird the only kind of free will worth wanting; incompatibilists disagree.

Who's right, metaphysically speaking, isn't our present concern. We want to do a better job of getting at what people ordinarily think. As we've seen, the extant data are mixed, but we don't think any of the previous studies have done a good enough job highlighting the issue we are emphasizing here – that is, the distinction between the kind of libertarian free will that one can exercise in the actual flow of events, keeping the past and the laws constant, and the merely conditional free will that is cashed out in terms of counterfactuals and what one could have done but only had the past and laws been different. According to our view, in a deterministic universe, once the past and the laws are fixed, everything we do is "in the cards," so to speak. At any given moment, we invariably do the one and only one thing that is open to us. Indeed, if we had a God's eye view, we would always be able to extrapolate what is going to happen at any given moment by looking at all of the antecedent circumstances. In this sense, there is no actual openness in a deterministic universe once the past and the laws are set. The only openness is merely counterfactual – that is, we have to first imagine the distal, antecedent circumstances having been different (otherwise, there is no room for a different present outcome).

The key issue, at least for us, is whether mere counterfactual possibilities – that is, the way things could have been had something else been different – are enough to ground the kind of free will most people believe in. Natural compatibilists think that for most people, merely having the conditional ability to do otherwise can ground free and responsibility agency. We have our doubts. We think most people share our sense that the kind of agency associated with compatibilism – whereby one cannot do otherwise in the actual stream of events, keeping the past and the laws constant – isn't enough to ground free will. Obviously, this is an empirical matter. Our goal is to shed some new light on the issue.

We therefore designed some vignettes that highlight precisely this issue (see below for details). Our sense is that compatibilists won't like the talk about "fixity," "causally closed (or open)," "a train going down the tracks," "counterfactual possibility and free will," "being able to know in advance what will happen," and the like.[12] However, we take all of these things to follow straightforwardly from the standard definition of determinism. In

a deterministic universe, given the past and the laws as they actually are at a particular place and time (and not merely as they could have been had things been different), the future is fixed, the universe is causally closed, everything is in the cards, and if one had a God's eye view, one could know everything that will ever happen in the future. In a deterministic universe, once the past and laws are fixed, the future is fixed – which is why it makes sense to talk about "fixity," "causally closed," and so on. The mere fact that the past and the laws could have been different in a deterministic universe doesn't make it any less fixed or closed, now that the past and laws are set. A deterministic universe is only open in a counterfactual sense. However, given that in the actual stream of events the past and the laws are actually fixed in a deterministic universe, there is one and only one way for things to unfold. That things could have been different only if other things had been different is no comfort to the agent doing the one and only thing open to him at the time. It is in precisely this sense that, metaphysically speaking, the way things unfold in a deterministic universe is like a "train going down the tracks."

While it is true that a deterministic universe is epistemically open – that is, the agent may not know where the train is invariably going – such a universe will unfold in the one and only one way it can, given the past and the laws. Even though such an agent may believe and feel as if more than one thing is genuinely open to him at the time, because a deterministic universe is epistemically open while metaphysically closed, these beliefs and feelings are illusory. Therefore, while compatibilists are sure to bemoan our wording and charge us with begging the question, we politely disagree. We made concrete (preregistered) predictions based on our intuition that the precise wording of our scenarios would clearly highlight the difference between the conditional and unconditional abilities to do otherwise – which we expected in turn to influence participants' intuitions about free will and responsibility in reliable ways. Keep in mind, we are not debating whether or not the conditional or the unconditional reading is the right metaphysical reading of the ability to do otherwise. We are debating whether the way we have operationalized these two abilities does justice to the difference between the two given the standard definition of determinism. We stand by our claim that it does.

Another worry likely to be raised by compatibilists is not that our wording begs the question against compatibilism, but rather, that our wording conflates determinism with fatalism. However, while there is an important metaphysical difference between the two,[13] we don't think we are conflating the two in our studies since we're careful to make sure that the fixity of the present is contingent on the fixity of the past. Because we explicitly build the conditional ability to do otherwise into our deterministic scenarios and make it clear that things could have been different had the

antecedents been different, we are not conflating determinism with fatalism. After all, the conditional ability to do otherwise is consistent with determinism, but inconsistent with fatalism. Therefore, we do not believe we are unfairly stacking the deck against compatibilism; indeed, quite the contrary. We think too many of the vignettes that have been used in the past have stacked the deck against incompatibilism by soft-pedaling the consequences of taking determinism seriously. For us, the key intuition that is relevant to the debate about natural compatibilism is whether people think that we can be free and responsible, even if, at every moment, we only have one and only one action available to us. We therefore decided to probe people's intuitions about the merely conditional ability to do otherwise by running three preregistered, between-subject, vignette-based studies.

3.3 Study 1: indeterminism versus determinism

3.3.1. Participants and design

For starters, in presenting and discussing Studies 1–3, we follow best scientific practices by reporting "how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study" (Simmons et al., 2012, p. 4). The complete data sets and supplemental materials (including measures and stimuli for all vignettes and further analyses that we did not have space to include in this paper) for all three studies can be found on our OSF page (https://osf.io/js8fa/). That said, we estimated the required sample size for Study 1 (as well as Studies 2 and 3) using power analysis in MATLAB (2018b) based on a preliminary study we ran that had two conditions (Determinism & Other and Indeterminism & Other) with 52 and 49 participants, respectively.[14] We excluded participants who failed to pass five comprehension questions from further analysis, which led to 17 and 21 participants for each condition. We predetermined a sample size required to achieve adequate power ($1 - \beta > 0.90$) for a two-sample t test, comparing the mean ratings of three free-will-related measures and the mean ratings of two moral-responsibility-related measures across two conditions. Given the effect sizes in this preliminary study, the power analysis indicated that we needed only 5 participants for each condition. However, because there was a large proportion of participants who failed comprehension questions, we wanted to be conservative and make sure that we would have a reasonable sample size across conditions. Therefore, we selected the sample size of 75 for each condition for all three studies, thinking this would give us more than enough power. We thought this struck an appropriate balance between having sufficient power to detect an effect and funding constraints.

For Study 1, data collection was stopped on the day that the minimum number of participants started the study. 349 individuals started (and 285

individuals completed) this two-by-two (Indeterminism vs. Determinism by Self vs. Other) between-subject study on Amazon's Mechanical Turk (MTurk), for monetary compensation.[15] Participant recruitment was restricted to individuals in the United States who had at least 1,000 previously accepted HITs (Human Intelligence Tasks) and a prior approval rating of at least 98%. 28 participants failed to follow instructions or failed to pass at least two out of three comprehension check questions, so data were analyzed with the 257 individuals (M_{age} = 42.76 years, SD = 13.56, $range_{age}$ = [21, 80], 123 females). Each individual was randomly assigned to one of the four conditions. All studies reported herein were approved by the College of Charleston Institutional Review Board.

For this vignette-based study, there were four conditions: (a) Indeterminism & Other, (b) Indeterminism & Self, (c) Determinism & Other, and (d) Determinism & Self. We used a between-subject design, so each participant received one and only one condition. For the other (or observer) conditions, the vignettes were as follows:

> **Indeterminism & Other condition**: Imagine Jim lives in a causally open universe. In this universe, despite the physical state of the universe, the laws of the universe, and the fixity of the past, at any given moment the universe is genuinely open, like a garden of forking paths. Whenever Jim makes a decision to act in a particular way, it's always the case that he could have acted differently even if absolutely everything leading up to his decision had been exactly the same. In short, at every moment Jim is able to actualize one possibility rather than another. Moreover, even if you knew absolutely everything about both the history of the universe and about Jim, you could never know in advance what Jim is going to decide to do. He, alone, is the only deciding factor when it comes to what he does. Despite the way the world was long before Jim was born, nothing in his life is in the cards, so to speak. Now, for illustrative purposes, imagine that Jim decided to take his dog for a walk in the park.
>
> **Determinism & Other condition**: Imagine Jim lives in a causally closed universe. In this universe, given the physical state of the universe, the laws of the universe, and the fixity of the past, at any given moment the universe is closed, like a train moving down the tracks. Whenever Jim makes a decision to act in a particular way, it's always the case that he could have acted differently only if something leading up to his decision had been different. In short, at any given moment, there is one and only one choice and action genuinely open to Jim. Moreover, if you knew absolutely everything about both the history of the universe and about Jim, you could always know in advance what Jim is going to decide to do. He is not the only deciding factor when it comes to what he does. Given the way the world was long before Jim was born, everything in his life is in the cards, so to speak. Jim can make choices, but these choices are the only choices open to him. Now, for illustrative purposes, imagine that Jim decides to take his dog for a walk in the park.

After reading one of these vignettes, participants in these conditions responded to the following statements on a 7-point scale ranging from *strongly disagree* to *strongly agree*:

(1) Jim has free will.
(2) Jim is in complete control of his choices and decisions.
(3) Jim's choices and decisions make a difference in what he does.
(4) In this scenario, free will is an illusion.[16]
(5) Jim is ultimately responsible for his actions.
(6) Jim is morally blameworthy for his bad actions and praiseworthy for his good actions.
(7) Jim could have decided not to walk his dog even if everything leading up to his decision remained the same.
(8) Jim could have decided not to walk his dog only if something leading up to his decision had been different.
(9) If one knew everything about Jim and the history of the universe, one could have known in advance that Jim was going to decide to walk his dog in the park.
(10) Everything Jim decides to do has to happen precisely as it does, given the state of the universe at the time of his decision.
(11) The universe Jim lives in is deterministic.
(12) The universe Jim lives in is indeterministic.

Items 1 to 4 were taken to elicit free-will judgments. Items 5 to 6 were taken to elicit moral-responsibility judgments. Items 7, 9, and 11 were used as comprehension checks.[17] The comprehension checks are condition-relative – for example, participants in the deterministic conditions (but not in the indeterministic conditions) needed to agree that the universe is deterministic, and participants in the indeterministic conditions (but not the deterministic conditions) needed to disagree that one could predict the protagonist's behavior in advance if one knew everything about the history of the universe. The items were presented in a random order to avoid framing effects. Agents who were in the self (or actor) conditions received these same vignettes, except that "Jim" was replaced by "you." The follow-up statements were reworded as well – for example, "in this scenario, you have free will," and "in this scenario, free will is an illusion."[18] After reading the vignette and responding to the 12 items, participants responded to the Free Will Inventory (FWI) (Nadelhoffer et al., 2014)[19] and provided some basic demographic information – for example, age, race, gender, education, income, political ideology, and religious orientation.

3.3.2. Results

To examine the effect of indeterministic and deterministic frames, and self and other perspectives on intuitions about free will and moral responsibility, as well as their interaction, we first aggregated and averaged the four free will questions (Cronbach's $\alpha = 0.91$)[20] and the two moral responsibility questions (Cronbach's $\alpha = 0.82$) and then performed a two-way Analysis of

Variance (ANOVA) and multiple comparison based on these two averages for participants who passed at least two out of three comprehension questions. Two-way ANOVAs indicated a main effect of indeterministic and deterministic frames on intuitions about free will, $F(1, 253) = 489.87$, $p < 0.001$, $\eta^2 = 0.66$, and intuitions about moral responsibility, $F(1, 253) = 206.53$, $p < 0.001$,$\eta^2 = 0.45$, (Figure 1). We did not observe a main effect of self and other perspectives on intuitions about either free will, $F(1, 253) = 0.16$, $p = 0.69$, or moral responsibility, $F(1, 253) = 2.81$, $p = 0.09$. We also did not find an interaction between these two factors on free will, $F(1, 253) = 0.01$, $p = 0.93$, or moral responsibility, $F(1, 253) = 0.03$, $p = 0.86$. Results were similar when all subjects were included (see Section 5 in the supplemental materials).

We performed a multiple comparison on indeterministic–deterministic frames and self–other perspectives across conditions corrected by Dunn and Sidák's approach (Sidák, 1967). Regarding intuitions about free will, four pairs were significantly different from each other, $p < 0.001$, but not the pair of Indeterministic–Other and Indeterministic–Self conditions, $p = 1.00$, or the pair of Deterministic–Other and Deterministic–Self conditions, $p = 1.00$. Specifically, the protagonists in the indeterministic conditions were perceived to have more free will than those in deterministic conditions (see Table 1 for details). Similarly, we found the same results for moral responsibility. Four pairs were significantly different from each other, $p < 0.001$, but not the pair of Indeterministic–Other

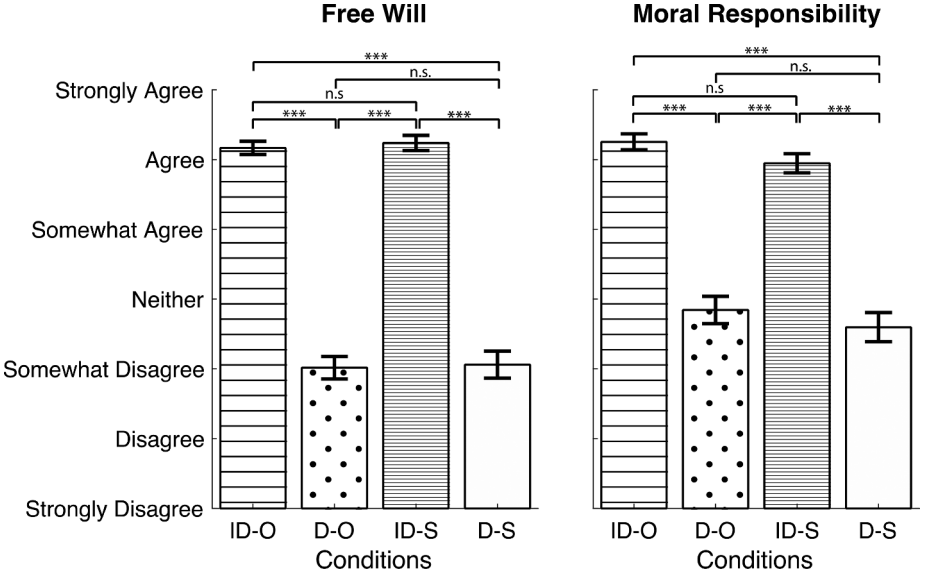

Figure 1. Aggregated free will and moral responsibility ratings. Error bars indicate the standard error from the mean. ID: indeterministic; D: deterministic; O: other or the third-person perspective; S: self or the first-person perspective. Note: n.s.: not significant; ***: $p < 0.001$.

Table 1. Statistics of free will and moral responsibility ratings.

	ID-Other	D-Other	ID-Self	D-Self
Free will	6.17 ± 0.79	3.02 ± 1.27	6.24 ± 0.88	3.06 ± 1.53
Moral responsibility	6.25 ± 0.95	3.84 ± 1.53	5.95 ± 1.10	3.60 ± 1.64

Note: each cell is denoted in Mean ± SD (standard deviation).

and Indeterministic–Self conditions, $p = 0.70$, or the pair of Deterministic–Other and Deterministic–Self conditions, $p = 0.88$. Specifically, attributions of moral responsibility were higher in the indeterministic conditions than in the deterministic conditions (Table 1). Results were similar when all subjects were included (see Section 5 in the Supplemental Materials).

Furthermore, contrary to our predictions, we did not find that attributions of moral responsibility were higher than attributions of free will across conditions, $t(512) = -1.67$, $p = 0.096$. Regarding Self–Other differences, in the deterministic scenario, we did not find higher self-attributions of free will, $t(121) = -0.17$, $p = 0.86$, or higher self-attributions of responsibility, $t(121) = 0.86$, $p = 0.39$.

3.3.3. Discussion

For Study 1, our goal was to compare intuitions in response to an indeterministic scenario which highlighted the unconditional ability to do otherwise with intuitions in response to a deterministic scenario which highlighted the merely conditional ability to do otherwise. We also used a morally neutral action – walking a dog in the park. Therefore, while the case was concrete, it was not affectively charged. Nevertheless, we found a stark difference between people's intuitions about free will and responsibility. In the indeterministic scenarios, people's responses averaged 6.2 in favor of free will and 6.1 in favor of moral responsibility. However, in the deterministic scenarios, people's responses were below the midpoint, averaging 3.04 for free will and 3.72 for moral responsibility.[21] Given that scores were below the midpoint (and not merely lower), these responses count in favor of natural incompatibilism. Keep in mind that we eliminated participants who incorrectly mislabeled the scenarios as indeterministic (when deterministic) or as deterministic (when indeterministic). We also made sure participants understood the nature of the unconditional or conditional nature of the agent's respective abilities. With these improved comprehension checks in place, we still found evidence for natural incompatibilism. These findings are in line with our initial predictions. However, much to our surprise, (a) there was only a marginal difference between intuitions about free will and intuitions about moral responsibility, and (b) people's self-attributions of free will and moral responsibility were not higher than their other-attributions.

3.4. Study 2: indeterminism versus neuro-determinism

3.4.1. Participants and experimental design

We once again predetermined a sample size of 75 for each condition for this study. Data collection was stopped on the day that the minimum number of participants started the study. 347 individuals started (and 290 individuals completed) this two-by-two (Indeterminism vs. Neuro-determinism by Self vs. Other) between-subject study on MTurk for monetary compensation. Participant recruitment was restricted to individuals in the United States who had at least 1,000 previously accepted HITs and a prior approval rating of at least 98%. 25 participants failed to follow instructions or failed to pass at least two out of three comprehension questions, so data were analyzed with the 265 individuals (M_{age} = 41.82 years, SD = 13.73, $range_{age}$ = [20, 80], 130 females). Each individual was randomly assigned to one of the four conditions. All studies reported herein were approved by the College of Charleston Institutional Review Board.

For this vignette-based study, there were four conditions: (a) Indeterminism & Other, (b) Indeterminism & Self, (c) Neuro-determinism & Other, and (d) Neuro-determinism & Self. We used a between-subject design, so each participant received one and only one condition. The indeterministic scenarios were the same as those used in Study 1. For the third-person neuro-deterministic conditions, participants read the following vignette:

> **Neuro-determinism & Other condition**: Imagine Jim lives in a causally closed universe. In this universe, given the physical state of the universe, the laws of the universe, and the fixity of the past, at any given moment the universe is closed, like a train moving down the tracks. Whenever Jim makes a decision to act in a particular way, it's always the case that he could have acted differently only if something leading up to his decision had been different. In short, at any given moment, there is one and only one choice and action genuinely open to Jim. Moreover, if you knew absolutely everything about both the history of the universe and about Jim, you could always know in advance what Jim is going to decide to do. He is not the only deciding factor when it comes to what he does. Given the way things were long before Jim was born, everything in his life is in the cards, so to speak. Jim can make choices, but these choices are the only choices open to him. These choices are the direct result of his past experiences, his present circumstances, and the current structural configuration of Jim's brain – which is like a complex biological computer. Indeed, Jim's choices and decisions are completely reducible to mechanistic neural events – which are just as causally closed as everything else in the universe. Now, for illustrative purposes, imagine that Jim decides to take his dog for a walk in the park.

The 12 statements were the same as those used in Study 1. We also followed the same procedure used in Study 1 for converting the vignettes and statements from third-person to first-person. After reading the vignette and responding to the 12 items, participants once again responded to the

Free Will Inventory (Nahmias et al., 2014) and provided some basic demographic information – for example, age, race, gender, education, income, political ideology, and religious orientation.

3.4.2. Results

To examine the effect of indeterministic and deterministic frames, and self and other perspectives on intuitions about free will and moral responsibility, as well as their interaction, we first aggregated and averaged the four free will questions (Cronbach's $\alpha = 0.90$) and the two moral responsibility questions (Cronbach's $\alpha = 0.86$), and then performed a two-way ANOVA and multiple comparisons based on these two averages of all participants who passed at least two out of three comprehension questions. Two-way ANOVAs indicated a main effect of indeterministic and deterministic frames on intuitions about free will, $F(1, 261) = 482.24$, $p < 0.001$, $\eta^2 = 0.65$, and intuitions about moral responsibility, $F(1, 261) = 192.30$, $p < 0.001$, $\eta^2 = 0.42$) (Figure 2). We did find a main effect of self and other perspectives on intuitions about moral responsibility, $F(1, 261) = 6.15$, $p = 0.014$, $\eta^2 = 0.013$, but not on intuitions about free will, $F(1, 261) = 0.05$, $p = 0.83$. We did not find an interaction between these two factors on free will, $F(1, 261) = 0.04$, $p = 0.84$, or moral responsibility, $F(1, 261) = 0.14$, $p = 0.71$. Results were similar when all subjects were included (see Section 5 in Supplemental Materials).

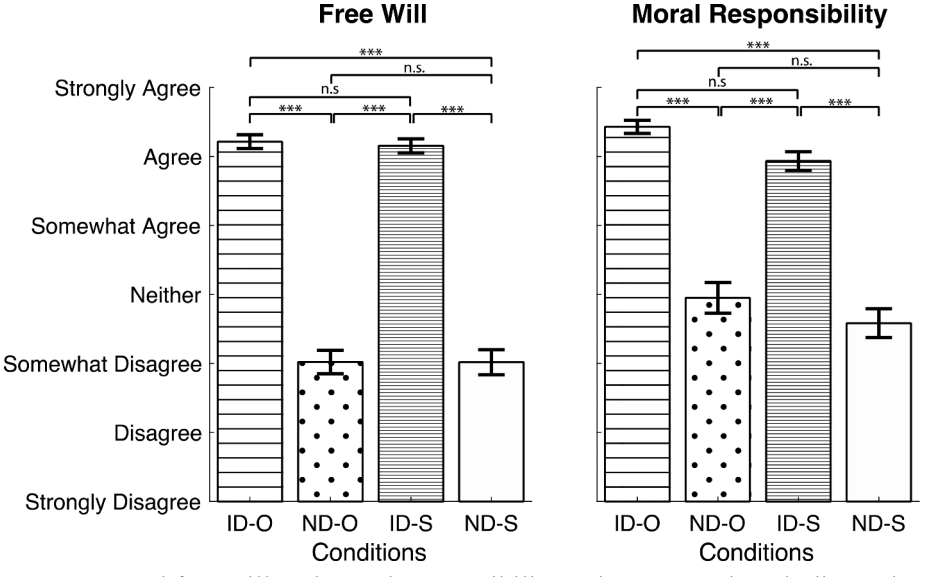

Figure 2. Aggregated free will and moral responsibility ratings. Error bars indicate the standard error from the mean. ID: indeterministic; ND: neuro-deterministic; O: other or the third-person perspective; S: self or the first-person perspective. Note: n.s.: not significant; ***: $p < 0.001$.

We performed multiple comparison on indeterministic–deterministic frames and self–other perspectives across conditions corrected by Dunn and Sidák's approach. Regarding intuitions about free will, four pairs were significantly different from each other, $p < 0.001$, but not the pair of Indeterministic–Other and Indeterministic–Self conditions, $p = 1.00$, or the pair of Deterministic–Other and Deterministic–Self conditions, $p = 1.00$. Specifically, the protagonists in indeterministic conditions were perceived to have more free will than those in deterministic conditions (see Table 2 for details). Similarly, we found the same results about moral responsibility. Four pairs were significantly different from each other, $p < 0.001$, but not the pair of Indeterministic–Other and Indeterministic–Self conditions, $p = 0.24$, or the pair of Deterministic–Other and Deterministic–Self conditions, $p = 0.56$. Specifically, attributions of moral responsibility were higher in the indeterministic conditions than in the deterministic conditions (Table 2). Results were similar when all subjects were included (see Section 5 in the Supplemental Materials).

Furthermore, this time around, unlike in Study 1, we found that attributions of moral responsibility were higher than attributions of free will across conditions, $t(528) = -3.42$, $p < 0.001$, $d = 0.19$. Regarding self–other differences, in the deterministic scenario, we did not find higher self-attributions of free will, $t(131) = 0.70$, $p = 0.61$, or moral responsibility, $t(131) = 1.20$, $p = 0.23$.

3.4.3. Discussion

For Study 2, our goal was to ramp up people's incompatibilist intuitions by using a case involving neuro-determinism – that is, a case that talks about the reductive nature of the brain and how it determines human thought and behavior. The question was whether we would once again find evidence for natural incompatibilism using a concrete case with a morally neutral action – walking a dog in the park. The responses in Study 2 were in line with Study 1. People who read the indeterministic scenario once again attributed free will and responsibility to the protagonist (whether the protagonist was them or someone else). The average ratings for these scenarios was 6.18 for free will, and 6.17 for responsibility. However, people who read the neuro-deterministic scenarios had lower attributions of free will (3.02) and responsibility (3.75).[22] Once again, these scores were below the midpoint, providing more evidence for natural incompatibilism. These findings

Table 2. Statistics of free will and moral responsibility ratings.

	ID-Other	ND-Other	ID-Self	ND-Self
Free will	6.21 ± 0.80	3.02 ± 1.32	6.15 ± 0.83	3.02 ± 1.53
Moral responsibility	6.43 ± 0.76	3.95 ± 1.74	5.93 ± 1.11	3.58 ± 1.77

Note: each cell is denoted in Mean ± SD (standard deviation).

comport with our prediction. The same can be said when it comes to the statistically significant difference in Study 2 between intuitions about free will and moral responsibility. As expected, the former was weaker than the latter. However, we once again did not find a difference between the responses to the first-person vignettes and the third-person vignettes. Given the literature on self–other differences when it comes to perceptions of agency and control, this is surprising. We're not sure what to make of our findings on this front.

3.5. Study 3: immoral indeterminism versus immoral neuro-determinism

3.5.1. Participants and experimental design

We once again predetermined a sample size of 75 for each condition for this study. Data collection was stopped on the day that the minimum number of participants started the study. 341 individuals started (and 281 individuals completed) this two-by-two (Indeterminism vs. Neuro-determinism by Self vs. Other) between-subject study on MTurk for monetary compensation. Participant recruitment was restricted to individuals in the United States who had at least 1,000 previously accepted HITs and a prior approval rating of at least 98%. 35 participants failed to follow instructions or failed to pass at least two out of three comprehension questions, so data were analyzed with the 246 individuals (M_{age} = 40.03 years, SD = 12.52, $range_{age}$ = [18, 72], 121 females). Each individual was randomly assigned to one of the four conditions. All studies reported herein were approved by the College of Charleston Institutional Review Board.

For this vignette-based study, there were four conditions: (a) Indeterminism & Other, (b) Indeterminism & Self, (c) Neuro-determinism & Other, and (d) Neuro-determinism & Self. We used a between-subject design, so each participant received one and only one condition. For Study 3, both the vignettes and the statements were the same as Study 2, except that rather than deciding to walk a dog in the park, the agent in the scenario (either "Jim" or "you") decided to murder a stranger. We once again used the same procedure for converting the third-person vignettes and statements into the first-person. After reading the vignette and responding to the 12 items, participants once again responded to the Free Will Inventory (Nahmias et al., 2014) and provided some basic demographic information – for example, age, race, gender, education, income, political ideology, and religious orientation.

3.5.2. Results

To examine the effect of indeterministic and deterministic frames, and self and other perspectives on intuitions about free will and moral responsibility, as well as their interaction, we first aggregated and averaged the four free

will questions (Cronbach's α = 0.90) and the two moral responsibility questions (Cronbach's α = 0.86), and then performed a two-way ANOVA and multiple comparison based on these two averages of all participants who passed at least two out of three comprehension questions. Two-way ANOVAs indicated a main effect of indeterministic and deterministic frames on intuitions about free will, $F(1, 242) = 442.82$, $p < 0.001$, $\eta^2 = 0.64$, and intuitions about moral responsibility, $F(1, 242) = 235.09$, $p < 0.001$, $\eta^2 = 0.47$ (Figure 3). We did find a main effect of self and other perspectives on intuitions about moral responsibility, $F(1, 242) = 13.19$, $p < 0.001$, $\eta^2 = 0.026$, but not on intuitions about free will, $F(1, 242) = 0.53$, $p = 0.47$. We found significant interaction between these two factors on free will, $F(1, 242) = 4.16$, $p = 0.043$,$\eta^2 = 0.006$, and moral responsibility, $F(1, 242) = 10.33$, $p = 0.0015$, $\eta^2 = 0.021$. Results were similar when all subjects were included (see Section 5 in the Supplemental Materials).

We performed multiple comparisons on indeterministic–deterministic frames and self–other perspectives across conditions corrected by Dunn and Sidák's approach. Regarding intuition about free will, four pairs were significantly different from each other, $p < 0.001$, but not the pair of Indeterministic–Other and Indeterministic–Self conditions, $p = 0.93$, or the pair of Deterministic–Other and Deterministic–Self conditions, $p = 0.27$. Specifically, attributions of free will to the protagonists in indeterministic conditions were lower than those in deterministic conditions (see

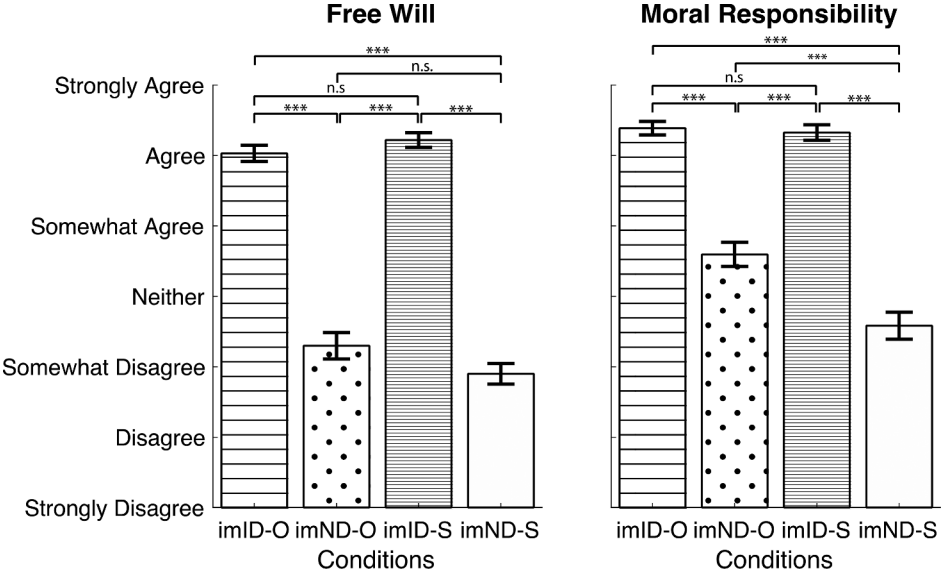

Figure 3. Aggregated free will and moral responsibility ratings. Error bars indicate the standard error from the mean. imID: immoral indeterministic; imND: immoral neuro-deterministic; O: other or the third-person perspective; S: self or the first-person perspective. Note: n.s.: not significant; ***: $p < 0.001$.

Table 3. Statistics of free will and moral responsibility ratings.

	imID-Other	imND-Other	imID-Self	imND-Self
Free will	6.03 ± 0.94	3.30 ± 1.49	6.22 ± 0.79	2.90 ± 1.14
Moral responsibility	6.39 ± 0.78	4.60 ± 1.36	6.32 ± 0.83	3.58 ± 1.49

Note: each cell is denoted in Mean ± SD (standard deviation).

Table 3 for more details). Similarly, we found the same results about moral responsibility. Five pairs were significantly different from each other, $p < 0.001$, but not the pair of Indeterministic–Other and Indeterministic–Self conditions, $p = 1.00$. Overall, the protagonists in indeterministic conditions were perceived to have more moral responsibility than those in deterministic conditions (Table 3). Results were similar when all subjects were included (see Section 5 in the Supplemental Materials).

Furthermore, we once again found that attributions of moral responsibility were higher than attributions of free will across all conditions, $t(490) = -3.87$, $p < 0.001$, $d = 0.35$. Regarding self–other differences, in the negatively valenced neuro-deterministic scenario, we did not find higher self-attributions of free will, $t(121) = 1.66$, $p = 0.10$, but we did find higher self-attributions of responsibility, $t(121) = 3.93$, $p < 0.001$, $d = 0.71$. In the negatively valenced indeterministic scenario, we did not find higher self-attributions of free will, $t(121) = -1.19$, $p = 0.23$, or higher self-attributions of responsibility, $t(121) = 0.43$, $p = 0.67$.

3.5.3. Discussion

For Study 3, we used the same vignettes from Study 2 except for one key difference – rather than using a morally neutral action (i.e., walking a dog), we used a morally charged action (i.e., murdering a stranger). Given that negatively valenced vignettes have been found in the past to elicit compatibilist responses, we wanted to see if we could offset the negative affect by focusing participants' attention on the difference between the unconditional and the conditional abilities to do otherwise. When it came to free will intuitions, that is precisely what we were able to do. In the indeterministic conditions, free will ratings averaged 6.12. In the deterministic conditions, on the other hand, the free will ratings averaged 3.1, once again crossing the midpoint, and once again providing evidence for natural incompatibilism.[23] Perhaps, unsurprisingly, intuitions about moral responsibility were stronger than free will intuitions. In the indeterministic scenarios, the responsibility ratings averaged 6.36, while in the deterministic scenarios, the responsibility ratings averaged 4.10. While the moral responsibility scores in the deterministic condition were much lower than in the indeterministic condition, they failed, just barely, to cross the midpoint. This suggests that people's intuitions about free will and moral responsibility come apart in negatively valenced scenarios – which might explain some of the earlier findings by

Nichols and Knobe (2007). As for self–other differences, while we did find a difference in one of the cases, as a general matter, we did not find the differences we expected from the outset. More research would need to be done to figure out why our predictions on this front failed.

4. Conclusions

There is an ongoing debate concerning the contours of people's beliefs about free will and moral responsibility. Evidence for both natural compatibilism and natural incompatibilism has been put forward. As such, researchers have found themselves at an empirical stalemate, with each party to the debate trying to explain away the conflicting data. Our hypothesis was that not enough attention had been paid to the metaphysically crucial difference between the unconditional and the conditional abilities to do otherwise. We therefore ran three preregistered between-subject vignette-based studies that focused on this issue. Across three studies involving concrete scenarios – including both morally neutral and negatively valenced actions – we found that people found determinism and neuro-determinism to undermine or challenge free will and responsibility. Given that we included improved comprehension checks to better ensure that participants were correctly understanding the indeterministic or deterministic nature of the scenarios (and the associated implications for agency), we believe that we have provided compelling evidence for natural incompatibilism.[24] While compatibilists will surely object to how we worded the scenarios, we were careful to motivate how these scenarios were designed. We believe these scenarios accurately describe some of the salient differences between determinism and indeterminism. As such, defenders of natural compatibilism have some further explaining to do.

One strategy that compatibilists may be inclined to adopt is to attempt to explain away our findings. According to this view, our deterministic stimuli may have primed what Eddy Nahmias has called "bypassing," that is, these stimuli illicitly induce participants to think that the agent's conscious beliefs, desires, and choices didn't make a difference in the agent's behavior (Nahmias, 2011; Murray & Nahmias, 2014; cf. Rose & Nichols, 2013). Because bypassing doesn't conceptually follow from determinism, this provides the compatibilist with an avenue for explaining away our findings, since they suggest some participants misunderstood the import of determinism. We anticipated this response, which is why we included an item that got at this worry, namely, "Jim's choices and decisions make a difference in what he does." Because this item had good reliability with the other free will items (which is telling), we aggregated them for the purposes of our main analyses. However, we also included the disaggregated analyses of the individual free will items (including the bypassing item) in

the Supplemental Materials, though this admittedly doesn't get at whether bypassing might partly explain our findings.

To shed light on whether our deterministic stimuli were inducing bypassing intuitions which, in turn, influenced participants intuitions about free will and responsibility, we first combined the data sets from our three studies. Then, following the method adopted by Murray and Nahmias (2014), we combined the free will items and the moral responsibility items (minus the aforementioned bypassing item) into a single variable. Finally, we ran mediation analyses to test two related models: (a) Determinism (D) versus Indeterminism (ID) > Bypassing (BP) > Free Will/Moral Responsibility (FW/MR), and (b) D versus ID > FW/MR > BP. The complete analyses for these two models can be found in Section 6 of the Supplemental Materials. Here, we just want to point out that Murray and Nahmias (2014) found that BP mediated the relationship between D versus ID and FW/MR, which they took to show that the deterministic stimuli were illicitly priming bypassing, which was, in turn, illicitly influencing ascriptions of free will and responsibility. However, the work by Rose and Nichols (2013) suggested that this conclusion was hasty. Their data suggested instead that FW/MR also mediates the relationship between D versus ID and bypassing, which is precisely what we found as well. Because mediation models only get at correlation and not causal directionality, Rose and Nichols (2013) went on to use structural equation modeling to shed light on the causal relationship between the variables, which revealed that deterministic stimuli influence ascriptions of free will, which, in turn, influence bypassing intuitions (and not the other way around). It seems that Murray, Nahmias, and others got the relationship backwards.

Unfortunately, given the number of dependent variables we used as part of our experimental design (since our focus was not on the issue of bypassing), our data aren't well-suited for structural equation modeling. However, given what Rose and Nichols (2013) found, and given that we, too, found that the relationship between bypassing intuitions and free will and responsibility intuitions was bidirectional, we think we have grounds for resisting the compatibilist's attempt to explain away our findings. While it is admittedly possible that our deterministic stimuli elicited bypassing intuitions – which influenced judgments about free will and responsibility, given both our mediation analyses and the work by Rose and Nichols (2013) – we think it is more likely that these stimuli elicited the opposite chain of judgments in a way that doesn't undermine the findings. Therefore, unless and until compatibilists explain away the findings by Rose and Nichols (2013), we think we're on relatively safe ground.

That said, our studies clearly have some limitations. First, we used exclusively online samples, and while, as we noted earlier, online data have been found to be just as reliable as – and more diverse than – data

collected on college campuses, we hope to extend our work in the future by collecting data from a convenience sample. Second, we only collected data from an American sample, so we can't generalize based on our findings. After all, our online participants were drawn from a country that is Western, Educated, Industrialized, Rich, and Democratic (WEIRD) (Henrich et al., 2010). While there are a limited number of cross-cultural studies that have explored free will beliefs (Hainnikainen et al., 2019; Sarkissian et al., 2010; Wisniewski et al., 2019), much work on this front remains to be done. In the meantime, it is worth noting that the bulk of the work on folk intuitions about free will has been done using WEIRD participants. Given that we were trying to respond to and build upon this research, it made sense for us to limit our attention to participants in the United States. However, we are quick to acknowledge that more cross-cultural work is required before we will know whether our findings are stable. Finally, while we used concrete cases involving both morally neutral and negatively valenced actions, it could be illuminating to explore abstract cases and also positively valenced actions. This is another logical extension of our work that we would obviously welcome.

Despite these limitations, we think we have advanced the debate concerning natural compatibilism by providing new evidence that people find free will and responsibility to be incompatible with determinism. When we made it clear to participants that determinism precluded the unconditional ability to do otherwise, and that indeterminism allowed for it, their judgments about free will and responsibility were influenced accordingly. We think that the gathering evidence now suggests that most people are indeterminists who associate free will with the unconditionality to do otherwise – an ability that all parties to the free will debate agree is incompatible with determinism. While more work clearly remains to be done, for now, we believe that we have shifted the empirical burden of proof squarely onto to the shoulders of the natural compatibilists who think that the conditional ability to do otherwise is supported by commonsense thinking about free will and moral responsibility.

Notes

1. In the wake of Harry Frankfurt's (1969) classic paper, "Alternative Possibilities and Moral Responsibility," some compatibilists have rejected the idea that free will even requires the conditional ability to do otherwise. Critics have suggested that Frankfurt's attempt to rid the free will debate of the so-called "principle of alternative possibilities" fails (e.g., Ekstrom, 2002; Franklin, 2011; Ginet, 1996; Van Inwagen, 1978; Widerker, 1995). Given the role that the ability to do otherwise has played – and continues to play – in the free will debate, for present purposes, we are going to set Frankfurt-style arguments aside. Therefore, when we talk about compatibilists in this

paper, we are focusing on the philosophers who think that the conditional ability to do otherwise is required for free and responsible agency.

2. For exceptions to this trend, see Deery et al. (2013), Deery et al. (2015), and Nahmias et al. (2004). See Section 2 for details.
3. There has been a lot of work in experimental philosophy on free will beliefs. We are only going to be able to focus on a narrow range of studies that are directly relevant to the task at hand.
4. The extant data make it clear that most people are pre-theoretical indeterminists (Bloom, 2012; Knobe, 2014; Turri, 2017). Consequently, participants often find descriptions of determinism to be counterintuitive if not implausible or even impossible (see Nahmias, 2006; Nahmias et al., 2005).
5. Rather than using the term 'determinism' – which is metaphysically loaded – Deery et al. (2013) introduced the notion of causal completeness: "according to causal completeness, everything that happens is fully caused by what happened before it. This is true from the very beginning of the universe, so that what happened in the beginning of the universe fully caused what happened next, and so on right up until the present. Causal completeness holds that everything is fully caused in this way, including people's decisions" (p. 133).
6. What Deery and colleagues say here is problematic. They make it sound as if the conflict method cannot shed light on the deliberative process involved in arriving at a judgment about case. After all, the only data point we have on this approach is the output. According to their view, "this makes it look as though all respondents took their answers to be obvious" (Deery et al., 2015, p. 778). However, we see no reason why this follows as a general rule. It's true that this is the case if researchers force participants to make dichotomous choices, however, as is now much more standardly the case, researchers ask whether participants agree or disagree – and to what extent. Thus, while some participants "somewhat agree," others "strongly agree." In this way, the conflict method is able to shed light on how obvious the judgments seemed to participants. This is not to suggest that the scale-based approach isn't a useful supplement, it's just to say that the criticisms of the conflict method by Deery and colleagues don't entirely hit the mark.
7. This partly undercuts the claim by Deery et al. (2015, p. 791) that their findings provide evidence that people are both incompatibilists and compatibilists. Until the two views are clearly parsed – which we think a number of their items arguably fail to do – this conclusion cannot be drawn. As is always the case, the devil is the details of the wording. Because we take issue with how a number of items are worded, we don't find their interpretation of their findings as compelling.
8. This problem was highlighted recently by the work on intrusive metaphysics by Rose et al. (2016).
9. For more on neuro-prediction and free will, see Deery et al. (2015), E. Nahmias et al. (2007), E. Nahmias et al. (2014), and Rose et al. (2016); Shepard & Reuter (2012).
10. We preregistered our studies both with AsPredicted and with the Open Science Framework (see https://osf.io/js8fa/).
11. We are not denying that counterfactual possibilities are real – they are every bit as real as actual possibilities. We are just highlighting that agents have two kinds of possibilities in a deterministic universe – the one and only actual option available to them at the time of choice, and the myriad different options that could have been open to them had things been different.
12. Someone might worry that we're using metaphorical language. However, we used metaphorical phrases because common language is often metaphorical, and since we

cannot use the terms 'determinism' and 'indeterminism,' we wanted to convey their meaning in a way that would be vivid to the participants. That philosophers use precisely these kinds of metaphors themselves when trying to illustrate the meaning of determinism and indeterminism suggests they are useful as rhetorical tools.

13. You can find earlier discussions of fatalism and its relevance to the debate about natural compatibilism in Feltz and Millan (2013), Nahmias (2006), and Nahmias and Murray (2011).
14. The dataset for this preliminary study is available on our OSF page, along with the rest of our materials.
15. Mturk is an online survey service that enables researchers to recruit and pay for participants for completing surveys of studies. For findings concerning the benefits of using MTurk – including the quality of the data and the improved diversity of the participant pool – see Burhmester et al. (2011), Paolacci et al. (2010), and Rand (2012).
16. For our main analyses for all three studies, which aggregate the four free-will-related items (i.e., free will, complete control, choices and decisions, and the illusion of control), this item was reverse-scored. You can find analyses of the disaggregated items in Section 4 of the Supplemental Materials.
17. We originally planned to use Items 7 to 12 as comprehension checks. However, upon further reflection, we noticed that Items 7 and 8, and Items 11 and 12 were each pairs of opposites. As such, we settled on 7 and 11, respectively, since we used a 7-point scale that allowed people to agree or disagree with each item. We also decided that Item 10 was too vague and hence unsuitable as a comprehension check. That left us with Items 7,9, and 11. We excluded all participants who missed at least two of these three comprehension checks. It turns out that excluding participants didn't make much of a difference when it comes to our central findings. See Section 5 of the Supplemental Materials for analyses of the three studies that don't include exclusions.
18. For all three studies, see Supplemental Materials for complete details.
19. Because we included the FWI as an exploratory measure but did not make any predictions concerning the results, we do not discuss the results here. We have, though, included the analyses of these findings from all three studies in Section 3 of the Supplemental Materials.
20. The disaggregated analyses of these four items can be found in Section 4 of the Supplemental Materials.
21. In the indeterministic scenarios, 5% disagreed with having free will (i.e., rated 1, 2, and 3), and 90% agreed (i.e., rated 5, 6, and 7); similarly, 5% disagreed with having moral responsibility (i.e., rated 1, 2, and 3), and 90% agreed (i.e., rated 5, 6, and 7). In the deterministic scenarios, 65% disagreed with having free will, and 22% agreed; whereas 44% disagreed with having moral responsibility and 37% agreed.
22. In the indeterministic scenarios, 5% disagreed with having free will (i.e., rated 1, 2, and 3), and 91% agreed (i.e., rated 5, 6, and 7); similarly, 2% disagreed with having moral responsibility (i.e., rated 1, 2, and 3), and 89% agreed (i.e., rated 5, 6, and 7). In the deterministic scenarios, 66% disagreed with having free will, and 25% agreed; whereas 45% disagreed with having moral responsibility and 38% agreed.
23. In the indeterministic scenarios, 6% disagreed with having free will (i.e., rated 1, 2, and 3), and 90% agreed (i.e., rated 5, 6, and 7); similarly, 1% disagreed with having moral responsibility (i.e., rated 1, 2, and 3), and 93% agreed (i.e., rated 5, 6, and 7). In the deterministic scenarios, 65% disagreed with having free will, and 24% agreed; whereas 39% disagreed with having moral responsibility and 46% agreed.

24. It is worth emphasizing that we are not suggesting that folk intuitions are univocal. Clearly, not all participants have the same intuitions about the cases – some give compatibilist answers, some give incompatibilist answers, and some neither agree nor disagree with the statements. As such, pluralism is the only way to adequately capture folk intuitions. However, just because pluralism is clearly the right view, we don't agree with Feltz et al. (2016) that this means that folk intuitions are irrelevant to issues like natural compatibilism. Even if intuition pluralism is true, we nevertheless think that our data speak to the truth of natural compatibilism – which is a majoritarian view. All that needs to true is that most people are naturally compatibilists (or incompatibilists). Therefore, just because not everyone has the same intuitions about free will, it doesn't mean these intuitions can't still be relevant to the debate about natural compatibilism.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Bloom, P. (2012). Free will does not exist. So what? *The Chronicle of Higher Education*. Retrieved May 22nd, 2019, from http://chronicle.com/article/Paul-Bloom/131170/

Burhmester, M., Kwang, T., & Gosling, S. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science*, *6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Clark, C. J., Lugari, J. B., Ditto, P. H., Knobe, J., Shariff, A. F., & Baumeister, R. F. (2014). Free to punish: A motivated account of free will belief. *Journal of Personality and Social Psychology*, *106*(4), 501–513. https://doi.org/10.1037/a0035880

Deery, O., Bedke, M., & Nichols, S. (2013). Phenomenal abilities: Incompatibilism and the experience of agency. In D. Shoemaker (Ed.), *Oxford studies in agency and responsibility* (pp. 126–150). Oxford University Press.

Deery, O., Davis, T., & Carey, J. (2015). The Free-Will Intuitions Scale and the question of natural compatibilism. *Philosophical Psychology*, *28*(6), 776–801. https://doi.org/10.1080/09515089.2014.893868

Ekstrom, L. (2002). Libertarianism and Frankfurt-style cases. In R. Kane (Ed.), *The Oxford handbook of free will* (pp. 309–323). Oxford University Press.

Feltz, A., Cokely, E. T., & Nelson, B. (2016). Experimental philosophy needs to matter: Reply to Andow and Cova. *Philosophical Psychology*, *29*(4), 567–569. https://doi.org/10.1080/09515089.2015.1125458

Feltz, A., & Millan, M. (2013). An error theory for compatibilist intuitions. *Philosophical Psychology*, *28*(4), 529–555. https://doi.org/10.1080/09515089.2013.865513

Frankfurt, H. (1969). Alternate possibilities and moral responsibility. *Journal of Philosophy*, *66*(23), 829–839. https://doi.org/10.2307/2023833

Franklin, C. (2011). Neo-Frankfurtians and buffer cases: The new challenge to the principle of alternative possibilities. *Philosophical Studies*, *152*(2), 189–207. https://doi.org/10.1007/s11098-009-9472-9

Ginet, C. (1996). In defense of the principle of alternative possibilities: Why I don't find Frankfurt's argument convincing. *Philosophical Perspectives*, *10*, 403–417.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Knobe, J. (2014). Free will and the scientific vision. In E. Machery & E. O'neill (Eds.), *Current controversies in experimental philosophy* (pp. 69–85). Routledge.

Lerner, M. J. (1980). *The belief in a just world: A fundamental delusion*. Plenum Press.

Murray, D., & Nahmias, E. (2014). Explaining away incompatibilist intuitions. *Philosophy and Phenomenological Research*, *88*(2), 434–467. https://doi.org/10.1111/j.1933-1592.2012.00609.x

Nadelhoffer, T., Shepard, J., Nahmias, E., Sripada, C., & Ross, L. (2014). The free will inventory: Measuring beliefs about agency and responsibility. Consciousness and Cognition, 25, 27–41 doi:10.1016/j.concog.2014.01.006

Nahmias, E. (2006). Folk fears about freedom and responsibility: Determinism vs. reductionism. *Journal of Cognition and Culture*, *6*(1–2), 215–237. https://doi.org/10.1163/156853706776931295

Nahmias, E., & Murray, D. (2011). Experimental philosophy on free will: An error theory for incompatibilist intuitions. In J. Aguilar, A. Buckareff, & K. Frankish (Eds.), *New waves in philosophy of action* (pp. 189–216). Palgrave-Macmillan.

Nahmias, E. (2011). Intuitions about free will, determinism, and bypassing. In R. Kane (Ed.), *The Oxford handbook of free will, 2nd Edition* (pp. pp. 555–576). Oxford University Press.

Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2004). The phenomenology of free will. *The Journal of Consciousness Studies*, *11*, 162–179.

Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2005). Surveying free will: Folk intuitions about free will and moral responsibility. *Philosophical Psychology*, *18*(5), 561–584. https://doi.org/10.1080/09515080500264180

Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2007). Is incompatibilism intuitive? *Philosophy and Phenomenological Research*, *73*(1), 28–53. https://doi.org/10.1111/j.1933-1592.2006.tb00603.x

Nahmias, E., Shepard, J., & Reuter, S. (2014). It's OK if 'my brain made me do it': People's intuitions about free will and neuroscientific prediction. *Cognition*, *133*(2), 502–516. https://doi.org/10.1016/j.cognition.2014.07.009

Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The Cognitive science of folk intuition. *Noûs*, *41*(4), 663–685. https://doi.org/10.1111/j.1468-0068.2007.00666.x

Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*, 411–419.

Pereboom, D. (2001). *Living without free will*. Cambridge University Press.

Rand, G. (2012). The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*, *299*, 172–179. https://doi.org/10.1016/j.jtbi.2011.03.004

Rose, D., Buckwalter, W., & Nichols, S. (2016). Neuroscientific prediction and the intrusion of intuitive metaphysics. *Cognitive Science*, *41*(2), 482–502. https://doi.org/10.1111/cogs.12310

Rose, D., & Nichols, S. (2013). The lesson of bypassing. *Review of Philosophy and Psychology*, *4*(4), 599–619. https://doi.org/10.1007/s13164-013-0154-3

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173–220). Academic Press.

Sarkissian, H., Chatterjee, A., De Brigard, F., Knobe, J., Nichols, S., & Sirker, S. (2010). Is belief in free will a cultural universal? *Mind & Language*, *35*(3), 346–358. https://doi.org/10.1111/j.1468-0017.2010.01393.x

Sidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, *62*(318), 626–633.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *Dialogue: The Official Newsletter of the Society for Personality and Social Psychology*, *26*, 4–7. https://doi.org/http://dx.doi.10.2139/ssrn.2160588

Taylor, S., & Brown, J. (1988). Illusion and well-being: A social psychological perspective of mental health. *Psychological Bulletin*, *103*(2), 193–210. https://doi.org/10.1037/0033-2909.103.2.193

Turri, J. (2017). Exceptionalist naturalism: Human agency and the causal order. *The Quarterly Journal of Experimental Psychology*, 1–16. https://doi.org/10.1080/17470218.2016.1251472

Van Inwagen, P. (1978). Ability and responsibility. *Philosophical Review*, *87*(2), 201–224. https://doi.org/10.2307/2184752

Widerker, D. (1995). Libertarianism and Frankfurt's attack on the principle of alternative Possibilities. *Philosophical Review*, *104*(2), 247–261. https://doi.org/10.2307/2185979

Wisniewski, D., Deutschländer, R., & Haynes, J-D. (2019). Free will beliefs are better predicted by dualism than determinism beliefs across different cultures. PLoS ONE, 14, e0221617. https://doi.org/10.1371/journal.pone.0221617