**Intuitions about Free Will and the Failure to Comprehend Determinism**

**Thomas Nadelhoffer, Samuel Murray, and Elise Dykhuis**

*Forthcoming in Erkenntnis*

**Introduction:**

Theories of free will are often partly measured against how well they accommodate relevant everyday intuitions. Frank Jackson, for instance, claims that the fundamental issue in the free will debate is ‘whether free action according to our ordinary conception, or something suitably close to our ordinary conception, exists and is compatible with determinism’ and he suggests that to identify our ordinary conception we must ‘appeal to what seems to us most obvious and central about free action … as revealed by our intuitions about possible cases’ (1998, p. 31). Others share Jackson’s views on the relevance of commonsense thinking to the free will debate. While incompatibilists like J.A. Cover and John O’Leary-Hawthorne claim that, ‘ordinary people have a conception of freedom, agency, and moral responsibility according to which an action is free and accountable only if that action is not fully determined by circumstances, past or present, that are beyond the agent’s control’ (1996, pp. 50-51), compatibilists like Daniel Dennett claim that when people ordinarily make attributions of moral responsibility, ‘it simply does not matter at all … whether the agent in question could have done otherwise in the circumstances’ (1984, p. 558).[[1]](#footnote-1)

Appeals to common sense are not merely rhetorical; rather, they shift the argumentative burden of proof onto those defending counter-intuitive views (Nahmias et al., 2005; 2006; Vargas, 2013). However, whether some theory is congruent with commonsense thinking about free will is an empirical issue. There is, after all, no reason to think that philosophers have special access to the contents of commonsense beliefs, free will included. Thus, for the past fifteen years, experimental philosophers have systematically explored intuitions about free will and the relationship between these intuitions and competing metaphysical theories of free will.[[2]](#footnote-2)

Empirical methods have found evidence for two different commonsense views of free will that correspond to the two predominant metaphysical theories of free will. Some evidence suggests that most people believe determinism precludes having free will, thereby indicating *natural incompatibilism* (see, e.g., Deery, Bedke, & Nichols, 2013; Nadelhoffer, Rose, Buckwalter, & Nichols, 2020; Nadelhoffer, Yin, & Graves, 2020; Nichols, 2006a; 2006b; 2012; Nichols & Knobe, 2007; Rose, Buckwalter, & Nichols, 2017; Roskies & Nichols, 2008). Other evidence suggests that most people do not believe determinism *per se* precludes having free will, thereby indicating *natural compatibilism* (see, e.g., Deery, Davis, & Carey, 2015 Murray & Nahmias, 2014; Nahmias, Morris, Nadelhoffer, & Turner, 2004; 2005; 2006; Nahmias & Murray, 2011; Woolfolk, Doris, & Darley, 2006).[[3]](#footnote-3) Both positions enjoy some empirical support. Hence, partisans of different metaphysical views of free will have undertaken to explain away empirical support for the opposing side.

Methodological problems, however, have recently stalled this experimental project. Mounting evidence suggests that a sizable percentage of participants fail to comprehend the nature and implications of determinism. In this paper, we explore the extent of this comprehension problem.[[4]](#footnote-4) We first set the stage with a survey of some key findings from the empirical literature. Then, we present the results of two preregistered studies designed to shed new light on the magnitude of the comprehension problem. Of the 585 participants who successfully completed our two studies, 81% (475) misinterpreted determinism by falsely believing either that determinism entails epiphenomenalism or fatalism or that determinism is compatible with fundamentally indeterministic capacities such as the unconditional ability to do otherwise. These results suggest that there is likely enough confusion concerning determinism baked into the existing evidence on folk intuitions about free will that these intuitions can’t reliably be used to shed light on the free will debate.

**Folk intuitions and the comprehension problem**

In some of the earliest work on folk intuitions about free will, Nahmias et al. (2005; 2006) found that a majority of participants (between 65%-85%) judged that agents in deterministic scenarios acted freely and were morally responsible. These findings were constant across scenarios 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). Nahmias and colleagues took their findings to provide compelling evidence for natural compatibilism.

Shortly thereafter, Nichols and Knobe (2007) published data on the psychological mechanisms that generate intuitions about moral responsibility. Participants were randomly assigned to read either abstract descriptions of both a deterministic Universe (A) and an indeterministic Universe (B) or similar descriptions that included concrete information about a person in Universe A, Bill, who murders his wife and family to be with his secretary.

In the concrete condition, 72% of participants gave the *compatibilist* response that Bill *is* fully morally responsible in Universe A (the deterministic universe). In the abstract condition, however, 84% gave the *incompatibilist* response that it is *not* possible for people in Universe A to be fully morally responsible. These findings challenged the claim that people’s intuitions are *robustly* compatibilist. Nichols and Knobe (2007) suggested that the compatibilist intuitions elicited in the concrete condition are best understood as performance errors driven by the affectively charged nature of the scenarios. Thus, they took their findings to provide compelling evidence for natural incompatibilism.

This has become a hallmark of the debate on folk intuitions about free will. Each side has developed error theories to explain away counter-evidence. For instance, some compatibilists argue that purportedly incompatibilist intuitions result from a belief in *epiphenomenal bypassing* (henceforth, *epiphenomenalism*), the thesis that in a deterministic system beliefs, desires, and other intentional mental states are causally inert (Murray & Nahmias, 2014; Nahmias 2011; Nahmias & Murray, 2011; cf. Björnsson, 2014; Björnsson & Pereboom, 2014; Rose & Nichols, 2013). Even incompatibilists agree that determinism does not entail epiphenomenalism (Kane, 2005, p. 19; van Inwagen, 1983, p. 23). So, because some intuitions about free will purportedly rest on this false belief about the inefficacy of mental states in deterministic systems, they are not properly incompatibilist intuitions, as they are grounded in a misunderstanding of the implications of determinism.[[5]](#footnote-5)

While some people seem to conflate determinism with epiphenomenalism, others seem to conflate determinism with fatalism (Nahmias and Murray, 2011).[[6]](#footnote-6) The latter comes in at least two varieties: On the one hand, there is what Andow and Cova (2016) have called ‘naïve fatalism,’ that is, the view that ‘whatever is going to happen, is going to happen, no matter what we do’ (Kane 2005, 19). On the other hand, there is the more modally sophisticated form of fatalism that Van Inwagen (1983) describes as ‘the thesis that it is a logical or conceptual truth that no one is able to act otherwise than he in fact does; that the very idea of an agent to whom alternative courses of action are open is self-contradictory’ (p. 3). While epiphenomenalism mistakenly assumes mental states are causally inert, fatalism mistakenly amplifies the modal strength of determinism. The thesis of determinism is that events occur necessarily conditional on some combination of laws of nature and past events. But determinism does not entail that there is only one logically possible set of laws and events (see, e.g., Audi, 1991, p. 320). Thus, it is a mistake to conflate □[(Po & L) → P] (determinism) with (□Po & □L) → □P (fatalism).[[7]](#footnote-7) Compatibilists, then, point to the false belief that determinism entails fatalism to explain away some purportedly incompatibilist intuitions.

Incompatibilists, on the other hand, argue that compatibilist intuitions may result from people importing indeterminacy into deterministic scenarios. Recent work has revealed that most people are natural indeterminists about human agency (Knobe, 2014; Monroe, Dillon, & Malle 2014; Turri, 2017). It is therefore unsurprising that indeterminism seems to influence people’s intuitions about free will in response to deterministic scenarios. For instance, researchers have found that people often mistakenly judge that *determinism* is compatible with possibilities that actually require *indeterminism*—e.g., the possibility that multiple outcomes are possible while holding fixed the past and the laws of nature (Nadelhoffer et al. 2020; Rose et al., 2017). These findings suggest that a sizable percentage of compatibilist intuitions are driven by an *intrusive indeterministic metaphysics* (henceforth, *intrusion*). This in turn suggests that these intuitions are not properly compatibilist since they, too, are predicated on a misunderstanding of determinism.

These dueling error theories pose a serious challenge: Some individual’s intuition that *p* lends evidential support to some theory that contains the statement that *p* *only if* the individual understands the theoretically relevant content of the proposition that *p*. If people systematically misinterpret the scenarios used to elicit intuitions that free will is (or is not) compatible with determinism, then these intuitions are theoretically irrelevant. Research on epiphenomenalism, fatalism, and intrusion suggests that a high percentage of participants from empirical studies misinterpret the scenarios by misunderstanding the nature of determinism depicted in them. However, the *depth* of the comprehension problem remains unknown, partly because there has been no effort to measure the three most common errors jointly.

1. **The present research**

We ran two preregistered studies to examine the extent to which participants misunderstand determinism. Our goal was to contribute to the ongoing debate in experimental philosophy about whether natural compatibilism or natural incompatibilism best capture folk intuitions about free will and moral responsibility. So, we selected two classic scenarios from Nichols and Knobe (2007) and Nahmias et al. (2005; 2006). We picked these scenarios because the former has been shown to elicit a high percentage of incompatibilist responses and the latter has been shown to elicit a high percentage of compatibilist responses. We wanted to explore the relationship between different sources of error and people’s intuitions about free will by using comprehension checks that enable us to detect participants who misunderstand the nature of determinism. This allowed us to assess the depth of the comprehension problem that we suspected has plagued existing research. Our findings suggest that the problem is much deeper than researchers have previously assumed.

* 1. *Study 1: Universe A vs. Universe B*
     1. *Participants*

In presenting and discussing Study 1 and Study 2, 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, Nelson, & Simonsohn, 2012, p. 4). All preregistrations, data sets, and supplemental materials (including measures and stimuli for all scenarios and further analyses excluded for space limitations) can be found at our OSF page.[[8]](#footnote-8) For Study 1, we recruited 377 participants[[9]](#footnote-9) from Amazon’s Mechanical Turk (MTurk).[[10]](#footnote-10) On average, it took participants 5.3 minutes to complete the study. They were paid $.60 for participating ($7.20/hour). 58 participants were excluded for not completing the survey, 19 participants were excluded for failing a ‘surface’ comprehension check (described below), and 8 were excluded because they claimed that Universe A is more like ours yet they also stated that Universe A is impossible (an incoherent pair of judgments). This left us with 292 participants (Mage = 36.8 years, SD = 10.7 rangeage: 21-70, 41% female, 58% Caucasian).[[11]](#footnote-11)

* + 1. *Methods and materials*

Study 1 used a between-subject design with three conditions: (a) epiphenomenalism, (b) fatalism, and (c) intrusion. All participants were first instructed to vividly imagine the following scenario from Nichols and Knobe (2007):

Imagine a universe (Universe A) in which everything that happens is completely caused by whatever happened before it. This is true from the very beginning of the universe, so what happened in the beginning of the universe caused what happened next, and so on right up until the present. For example one day John decided to have French Fries at lunch. Like everything else, this decision was completely caused by what happened before it. So, if everything in this universe was exactly the same up until John made his decision, then it had to happen that John would decide to have French Fries.

Now imagine a universe (Universe B) in which almost everything that happens is completely caused by whatever happened before it. The one exception is human decision-making. For example, one day Mary decided to have French Fries at lunch. Since a person’s decision in this universe is not completely caused by what happened before it, even if everything in the universe was exactly the same up until Mary made her decision, it did not have to happen that Mary would decide to have French Fries. She could have decided to have something diﬀerent.

The key diﬀerence, then, is that in Universe A every decision is completely caused by what happened before the decision— given the past, each decision has to happen the way that it does. By contrast, in Universe B, decisions are not completely caused by the past, and each human decision does not have to happen the way that it does.

Having read the scenario,[[12]](#footnote-12) participants were first asked which of these two universes is the most similar to our own (Universe A or Universe B)*.* Second, they were asked whether they thought that Universe A is possible (Yes or No). Third, they were presented with the following surface comprehension check:[[13]](#footnote-13) ‘In Universe A, everything that happens is completely caused by what happened before it’ (True of False).[[14]](#footnote-14) The people who missed this surface comprehension check were excluded from our analysis. Next, participants indicated how vividly they could imagine the scenario using a 5-pt. Likert scale (1 = very slightly or not at all to 5 = extremely)[[15]](#footnote-15) before reading the following prompt:

Regardless of how you answered the previous questions, imagine that it is possible that there is a universe where everything that happens is completely caused by what happened before it, such that, given the past, everything that happens has to happen. With that in mind, please read the following statements and note your level of agreement. We are only interested in your honest opinions. There are no right or wrong answers. So, please read the items very carefully and respond accordingly.[[16]](#footnote-16)

All participants then stated their level of agreement with these items using a 7-pt. Likert (1=strongly disagree to 7=strongly agree):

1. It is possible for a person in Universe A to have free will.[[17]](#footnote-17)
2. It is possible for a person in Universe A to be morally blameworthy when they do something wrong.

Depending on condition, participants also saw one of the following sets of statements:[[18]](#footnote-18)

Epiphenomenalism:[[19]](#footnote-19)

1. In Universe A, what people want and believe has no effect on what they do.
2. In Universe A, John would have decided to have French Fries no matter what he wanted or believed.
3. In Universe A, it doesn’t make any sense to say that John made his own choice to eat French Fries.
4. In Universe A, it’s not up to John whether or not to eat fries.

Fatalism:[[20]](#footnote-20)

1. In Universe A, there is no sense in which events could have unfolded differently than they did.
2. In Universe A, John would have ended up having French Fries no matter what he tried to do.
3. In Universe A, John will eat French Fries no matter what.
4. In Universe A, John’s eating French Fries had to happen, *even* if what happened in the past had been different.

Intrusion:

1. In Universe A, what people decide to do could have been different even if everything leading up to the decision had been exactly the same.
2. In Universe A, there was a slight chance that John could have chosen not to have French Fries at the time.
3. In Universe A, it was open for John to choose not to have French Fries at the exact moment he decided to have them.
4. In Universe A, John could have decided not to have French Fries even though his decision to have them was completely caused.

Finally, we asked participants whether they had taken a free will-related survey in the past[[21]](#footnote-21) and collected basic demographic information (race, age, gender, etc.).

Based on Nichols and Knobe’s (2007) results, we predicted that most people would think our universe is most like indeterministic Universe B and that they would disagree that a person in Universe A could have free will or be morally blameworthy. However, we also predicted that most people would judge that Universe A is impossible and that they would agree with the epiphenomenalism items, the fatalism items, or the intrusion items (depending on the condition)—which would indicate that these participants misunderstood some aspect of the deterministic nature of the scenario. Finally, we predicted that agreement with these comprehension items would correlate with intuitions about free will in predictable ways—e.g., people who agree with epiphenomenalism or fatalism will be less likely to attribute free will while people who agree with intrusion will be more likely to attribute free will.

* + 1. *Results and discussion*

Across all conditions, 150 participants rated Universe B (indeterministic) as most similar to our own, while 142 participants rated Universe A (deterministic) as most similar, and a one-sample proportion test found no statistically significant difference in the proportion of responses (c2(1) = 0.22, *p* = .64). This finding is quite different from results reported in Nichols and Knobe (2007), where 90% of participants judged that Universe B was more similar. Finally, more participants in our study rated Universe A as possible (223) rather than impossible (69), and a one-sample proportion test indicated this was a statistically significant difference (c2(1) = 81.2, *p* < .001).

Our results have two significant findings. First, we found that participants were pretty evenly split on intuitions about free will, with 50% (146 participants) judging that people could have free will in a deterministic universe and 42% (124 participants) judging that people could not have free will in such a universe.[[22]](#footnote-22) This is certainly not compelling evidence for natural incompatibilism. The responses to the moral responsibility item were even more surprising. While 86% of the participants from the study by Nichols and Knobe judged that people could not be fully morally responsible in a deterministic universe, we found instead that 58% (170) of our participants judged that people *could be* responsible whereas 30% (87) judged that people could not be morally responsible. These are markedly different results from Nichols and Knobe (2007). We found little evidence for natural incompatibilism in people’s judgments about free will and moral responsibility, and failed to replicate some of the findings from Nichols and Knobe.[[23]](#footnote-23) However, these differences might be attributable to different measures and better sampling. Whatever the reason, our primary concern was not replication, but checking for comprehension.

In our eyes, the most significant finding from Study 1 was the widespread failure to comprehend determinism. 98% of participants (95/97) in the first condition agreed with one or more statements that indicate mistaking determinism for epiphenomenalism, 95% of participants (96/101) in the second condition agreed with one or more statements that indicate mistaking determinism for fatalism, and 67% of participants (62/92) in the third condition agreed with one or more statements that indicate intrusion. The internal reliability was adequate for items measuring epiphenomenalism (a = .61) and fatalism (a = .60) and high for intrusion (a = .94). Cronbach’s alpha (a) is a measure of shared covariance (or consistency) between continuous variables. Thus, average scores were likely not distorted by a single outlier item.

However, consistency does not entail unidimensionality. Because the measures of epiphenomenalism, fatalism, and intrusion are new, we wanted to ensure that the items within each scale were mapping to the same construct. We ran exploratory factor analyses (EFA) for each scale, an analysis that shows how a set of items group together based on the covariance of the items. The results are in the Supplemental Material on our OSF page. For present purposes, we simply want to note that, combining responses across both Studies 1 and 2, we conducted three separate EFAs on the three scales. In each instance, the four items for a particular scale were entered for analysis, using the criterion of eigenvalues greater than 1 and maximum likelihood extraction with varimax rotation. The results in each analysis indicated that the four items measured one underlying construct, thus indicating three unidimensional measures, one for each construct. Taken together, the results of the EFAs, combined with the findings on reliability, should help allay concerns about how our individual items are worded and theoretically group together. Despite wording epiphenomenalism, fatalism, and intrusion in different ways, people’s responses to the items for each type of error were largely of a piece.

No matter how we worded the items, most participants mistakenly intuited that determinism entails either epiphenomenalism or fatalism or that it is compatible with capacities that require indeterminism—e.g., the unconditional ability to do otherwise. Keep in mind, our analysis only includes participants who *passed* the kind of surface comprehension check that has been used by researchers in this past as the *only* measure of comprehension. As such, our results clearly indicate that surface comprehension checks are insufficient when it comes to controlling for the comprehension problem. After all, most of the participants who passed a surface comprehension check in our first study went on to fail one or more fine-grained comprehension checks later in the study. In each of these latter cases, the participants misunderstood determinism.

Widespread misunderstanding of determinism is bad enough in itself, but it is made worse by the fact that it was related in predictable ways to participants’ intuitions about free will and responsibility (see **Table 1**).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Pearson’s correlations for judgments of free will and responsibility in Study 1 | | | | | | | |
|  | FW | MR |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| 1. Epiphenomenalism | -0.16 | -0.22\* |  |  |  |  |  |
| 2. Intrusion | 0.81\*\*\* | 0.70\*\*\* |  |  |  |  |  |
| 3. Fatalism | 0.00 | -0.11 |  |  |  |  |  |
| *Note*. Correlations are computed using mean response to the four items of the error scale.  \*\*\*p < .001; \*\*p < .01; \*p < .05 | | | | | | | |

Pearson’s product-moment correlation indicated that (a) epiphenomenalism was negatively correlated with moral responsibility intuitions (r = -0.22, *p* = .032, 95% *CI*[-0.40, -0.02]), (b) fatalism was not correlated with free will or responsibility intuitions, which is surprising,[[24]](#footnote-24) and (c) intrusion was strongly correlated with both free will intuitions (*r* = 0.81, *p* < .001, 95% *CI*[0.73, 0.87]) and moral responsibility intuitions (*r* = 0.70, *p* < .001, 95% *CI*[0.58, 0.79]) (see **Figure 1**). Epiphenomenalism and intrusion thus pull in *opposite* directions—the former is associated with weaker beliefs about agency while the latter is associated with stronger beliefs.



**Figure 1***.* Scatter plots for correlations between average free will and moral responsibility ratings and epiphenomenalism (Panel A), intrusion (Panel B), and fatalism (Panel C) in Study 1. Error bars represent 95% confidence intervals.

The results concerning intrusion in particular suggest an underlying strand of incompatibilism in commonsense thinking about free will since they highlight the association people draw between indeterminism, free will, and moral responsibility. Moreover, these results suggest that intrusion is a more pressing comprehension error than epiphenomenalism and fatalism, as intrusion errors are very strongly correlated with judgments about free will and moral responsibility—greatly inflating the appearance of seemingly compatibilist intuitions. That said, it is worth emphasizing that we purposely used a scenario in Study 1 that has previously been shown to elicit predominantly incompatibilist intuitions. So, for Study 2, we used a scenario that elicits predominantly compatibilist intuitions. The results are just as worrisome.

* 1. *Study 2: Supercomputer*
     1. *Participants*

For Study 2, we recruited 325 participants from MTurk (based on the samepower analysis used in Study 1). On average, it took participants 4.9 minutes to complete the study. They were paid $.60 for participating ($7.34/hour). 30 participants were excluded for not completing the survey, 16 were excluded for failing a surface comprehension check (described below), and 7 were excluded because they claimed both that the universe in the scenario is similar to our own and that it is impossible (which is incoherent). This left us with 272 participants (Mage = 37.7 years, SD = 10.8 rangeage = 22-77, 41.4% female, 72.4% Caucasian).[[25]](#footnote-25)

* + 1. *Methods and materials*

Study 2 utilized the same design as Study 1 except we used a scenario that elicits a high percentage of compatibilist responses to balance out the incompatibilist-leaning scenario from Study 1. So, we told participants to vividly imagine the following modified version of the Supercomputer scenario from Nahmias et al. (2005; 2006):[[26]](#footnote-26)

Imagine a Universe where research scientists discover all the laws of nature and build a supercomputer which can deduce from these laws of nature and from the current state of everything in the world exactly what will be happening in the world at any future time. It can look at everything about the way the world is and predict everything about how it will be with 100% accuracy. Suppose that in this Universe, the supercomputer looks at the current state of things and deduces that a certain individual, Jeremy Hall, will definitely rob a bank at 6:00PM in 45 years. As always, the supercomputer’s prediction is correct: Jeremy robs the bank at 6:00PM on the exact date predicted by the computer.

Having read the scenario, participants were asked how similar they think this universe is to our own universe (ranging from very dissimilar to very similar)*.* Second, they were asked whether they thought that this universe is possible (Yes or No). Third, they were presented with the following surface comprehension check: ‘According to the scenario, the supercomputer can deduce from the laws of nature and from the current state of everything in the world exactly what will be happening in the world at any future time’ (True or False).[[27]](#footnote-27) As before, participants who missed this surface comprehension check were excluded from our analysis. Next, participants indicated how vividly they could imagine the scenario using a 5-pt. Likert scale (1 = very slightly or not at all, 5 = extremely) before reading the following prompt:

Regardless of how you answered the previous questions, imagine such a supercomputer actually did exist and actually could perfectly predict the future, including Jeremy’s robbing the bank (and assume Jeremy does not know about the prediction). With that in mind, please read the following statements and note your level of agreement. We are only interested in your honest opinions. There are no right or wrong answers. So, please read the items very carefully and respond accordingly.

All participants then stated their level of agreement with these items using a 7-pt. Likert (1=strongly disagree to 7=strongly agree):

1. Jeremy robbed the bank of his own free will.
2. Jeremy is morally blameworthy for robbing the bank.
3. Jeremy’s robbing the bank was a bad thing.[[28]](#footnote-28)

Depending on condition, participants also saw one of the following sets of items:

Epiphenomenalism:

1. What Jeremy wanted and believed had no effect on whether he robbed the bank.

2. Jeremy would have decided to rob the bank no matter what he wanted or believed.

3. It doesn’t make any sense to say that Jeremy made his own choice to rob the bank.

4. It’s not up to Jeremy whether or not to rob the bank.

Fatalism:

1. Jeremy could have decided not to rob the bank even if everything (including the laws of nature) had been exactly the same prior to his decision.
2. There was at least a slight chance that Jeremy could have chosen not to rob the bank even if everything (including the laws of nature) had been exactly the same prior to his decision.
3. It was open for Jeremy to choose not to rob the bank at the exact moment he decided to rob it.
4. Jeremy could have decided not to rob the bank even though the computer predicted he would rob the bank based on the state of the universe and the laws of nature.

Intrusion:

1. There is no sense in which events could have unfolded differently than they did.
2. Jeremy would have ended up robbing the bank no matter what he tried to do.
3. Jeremy will rob the bank no matter what.
4. Jeremy’s robbing the bank had to happen, even if what happened in the past had been different.

Finally, we asked participants whether they had taken a free will-related survey in the past and collected basic demographic information (race, age, gender, etc.).

We predicted that most people would attribute free will and moral blameworthiness to the agent in the scenario (thereby replicating Nahmias et al.). Second, we expected that, across all conditions, most participants would misunderstand determinism. Finally, based on results of a pilot study, we expected agreement with epiphenomenalism and fatalism would be lower in Study 2 than in Study 1, while agreement with intrusion would be higher in Study 2—which would show that while comprehension failures are widespread, different ways of misunderstanding determinism are more or less likely to be elicited depending on the wording of the scenario.

* + 1. *Results and discussion*

Because the Supercomputer scenario does not describe two universes, we measured how similar the Supercomputer universe seemed to our own. Participants saw the Supercomputer universe as relatively dissimilar to our own, with no significant differences across conditions (*M* = 1.89, *SD* = 1.73, *F*(2, 292) = 0.22, *p* =0.80, h2p = .001, 95% *CI* [0.00, 0.02]). Finally, while a greater number of participants believed the Supercomputer universe was impossible (158) rather than possible (137), this difference was not significant (c2(1) = 1.49, *p* = 0.22).

As predicted, the overwhelming majority of participants judged that Jeremy has free will (76%, 223 participants) and that he is morally blameworthy (80%, 237 participants).[[29]](#footnote-29) This replicates the results of Nahmias and colleagues. Determinism seems less threatening to people’s beliefs about free will and responsibility in the Supercomputer scenario relative to the Universe A/B scenario. However, responses to the comprehension checks present a possible deflationary explanation of this difference.

As predicted, agreement with epiphenomenalism and fatalism was higher in Study 1 (epiphenomenalism *M* = 5.71; fatalism *M* = 5.48) relative to Study 2 (epiphenomenalism *M* = 3.91; fatalism *M* = 4.68), with a medium to large effects of Vignette on average responses (epiphenomenalism: *t*(152.6) = 8.92, Welch’s *p* < .001, *d* = 1.29, 95% *CI*[0.98, 1.60]; fatalism: *t*(192.11) = 4.58, Welch’s *p* < .001, *d* = 0.64, 95% *CI*[0.35, 0.92]), which would explain the disparity in judgments of free will and moral responsibility between the two studies. However, the comprehension failure rate in Study 2 was quite high. We also found significantly higher agreement with intrusion in Study 2 (Study 1 *M* = 4.04; Study 2 *M* = 4.82), and a two-tailed *t*-test found a medium effect of Vignette on average intrusion responses (*t*(184.12) = -2.79, Welch’s *p* = .006, *d* = 0.41, 95% *CI*[0.12, 0.70]). This might also explain the disparity in judgments between the two. The materials in Study 1 prompt attitudes that are associated with diminished judgments of freedom and responsibility, while the materials in Study 2 prompt attitudes that are associated with amplified judgments of freedom and responsibility. The three error measures also exhibited better internal reliability in Study 2 for epiphenomenalism (a = .87), fatalism (a = .87), and intrusion (a = .95).[[30]](#footnote-30)

As in Study 1, misinterpretations of determinism correlate predictably with intuitions about free will and responsibility (see **Table 2**).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Pearson’s correlations for judgments of free will and responsibility in Study 1 | | | | | | | |
|  | FW | MR |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| 1. Epiphenomenalism | -0.53\*\*\* | -0.49\*\*\* |  |  |  |  |  |
| 2. Intrusion | 0.81\*\*\* | 0.70\*\*\* |  |  |  |  |  |
| 3. Fatalism | 0.00 | 0.11- |  |  |  |  |  |
| *Note*. Correlations are computed using mean response to the four items of the error scale.  \*\*\*p < .001; \*\*p < .01; \*p < .05 | | | | | | | |

We found (a) epiphenomenalism was negatively correlated with free will intuitions (*r* = -0.53, *p* <. 001, 95% *CI*[-0.66, -0.37]) and moral responsibility intuitions (*r* = -0.49, *p* <. 001, 95% *CI*[-0.63, -0.32]), (b) fatalism was not correlated with free will (*r* = -0.14, *p* = 0.16, 95% *CI*[-0.33, 0.05] or responsibility intuitions (*r* = -0.01, *p* = 0.92, 95% *CI*[-0.20, 0.18]), and (c) intrusion was positively correlated with both free will intuitions (*r* = .610, *p* < .001, 95% *CI*[0.47, 0.72]) and moral responsibility intuitions (*r* = 0.54, *p* < .001, 95% *CI*[0.38, 0.67]) (see **Figure 2** for summary).



**Figure 2***.* Scatter plots for correlations between average free will and moral responsibility ratings and epiphenomenalism (Panel A), intrusion (Panel B), and fatalism (Panel C) in Study 2. Error bars represent 95% confidence intervals.

These results do not bode well for researchers who want to use the findings from the Supercomputer scenario to support natural compatibilism. Considered against the backdrop of what we found in Study 1, these results indicate that a serious comprehension problem lies at the root of experimental studies of everyday thinking about free will. Treating the epiphenomenalism, fatalism, and intrusion items as proper comprehension checks would require us to exclude over 80% of our participants (depending on both the scenario and the specific comprehension failure).[[31]](#footnote-31)

This is particularly alarming since we think that we *should* treat these items as comprehension checks. If participants think that determinism entails epiphenomenalism or fatalism, or that determinism is compatible with capacities that require indeterminism, then they have clearly misunderstood the nature and implications of determinism as philosophers understand it. If that’s right, then the extant data, insofar as it relies on these measures and scenarios, do not speak to whether or not commonsense thinking about free will reflects compatibilist or incompatibilist commitments.

1. **General discussion**

There is a long tradition of aligning one’s theory of free will with common sense. Experimental philosophers have tried to shed empirical light on the commitments of commonsense thinking about free will. These efforts have relied on the largely unchecked assumption that people understand the concepts used to elicit the relevant intuitions. As we have seen, independent lines of research recently began investigating whether people understood the stimuli used in free will studies. Our studies, however, represent the first attempt to study these three comprehension errors in tandem while using four items for measuring each error. The results are striking for at least two reasons. First, comprehension of basic deterministic elements was very low. Second, we only checked for three comprehension errors. It is possible that, had we checked for more, nearly all of our participants could have been excluded for misunderstanding determinism. Importantly, we analyzed data from only those participants who passed standard comprehension checks. Because passing such checks requires some degree of attentiveness to details of the materials, it is likely *not* the case that the comprehension errors we observed can be explained entirely by carelessness or distracted responding. In this respect, passing the surface comprehension check functions as a measure of “good faith” responding.

Our work highlights a general problem running through experimental work on commonsense thinking about free will. The traditional debate about free will focuses on whether it is incompatible with determinism *as philosophers understand it*. In order for data on folk intuitions to be probative to whether common sense contains incompatibilist or compatibilist elements, participants must understand determinism in roughly the same way as philosophers do. All parties to the philosophical debate agree that determinism does not entail epiphenomenalism or fatalism and that determinism is incompatible with capacities like the unconditional ability to do otherwise. Our findings suggest that this is not how most participants understand determinism.

These remarks provide an important qualification on what it means to say that participants ‘misunderstand’ or commit ‘comprehension errors’ with respect to determinism. Our results show that people are not representing facts about the kind of determinism at issue in philosophical debates about free will. It is in this respect that participants exhibit misunderstanding about determinism. This does not imply that how participants interpret our materials—or how participants ordinarily think of determinism—is irrelevant. However, insofar as one is trying to use folk intuitions about determinism as evidence for the congruence between a theoretical characterization of free will and everyday thinking about it, our results suggest that standard research materials and designs have not elicited folk judgments that are theoretically probative. While some might consider finding such congruence superfluous, others (mentioned in the Introduction) have found it important to align theory with practice. Within that context our results can be summarized as suggesting that current empirical methods have not yet understood the *practice* side of thinking about free will, so there is no clear way for it to inform the *theory* side of thinking about free will. The ‘errors’ we attribute to participants are nothing more than mistaken characterizations of philosophical constructs that disrupt the connection between practice and theory.[[32]](#footnote-32)

Even though we primarily measured participant errors, we think that some of our results can be pressed into theoretically interesting service. For example, some of our findings have ramifications for views about the structure of the folk psychological concept of free will. Many views of free will assume *invariantism*, or the thesis that people represent a basic set of criteria for making judgments about free will. Recently, however, some have argued for *variantism*, or the thesis that people represent distinct (mutually incompatible criteria) in different situations when making judgments (Knobe & Doris, 2010; Nichols, 2015). Empirical evidence plays a significant role in the argument for variantism. Knobe and Doris, for example, say the following: ‘People’s ordinary responsibility judgments do not appear to follow from a single system of criteria that can be applied in all cases’ (p. 346). The ordinary judgments are measured in experiments like the ones used here, relying on vignettes to elicit intuitions about who is free and responsible and to what degree. Notably, both the Universe A/B and Supercomputer scenarios figure prominently in building a case for variantism. Our results suggest that the evidence for variantism is lacking because the empirical study of free will beliefs tells us little about the criteria people use for making judgments about free will. This is because people systematically misinterpret the cases used to elicit such judgments. In other words, people’s pervasive mistaken beliefs about determinism hopelessly confound the empirical evidence.[[33]](#footnote-33)

Our studies rely on materials that represent specific characterizations of determinism. The Universe A/B vignette describes determinism in terms of the past completely causing states in the present, while the Supercomputer vignette describes determinism in terms of a machine capable of perfectly predicting any point in the future based on complete knowledge of the present. These might seem to be inadequate characterizations of determinism. For example, some philosophers argue that determinism can be defined independently of the notion of causation, meaning that determinism is not equivalent to the past completely causing the present (e.g., van Inwagen, 1983, pp. 3-4). Instead, determinism is a thesis about entailment:

For every instant of time, there is a proposition that expresses the state of the world at that instant; If *p* and *q* are any propositions that express the state of the world at some instants, then the conjunction of *p* with the laws of nature entails *q* (van Inwagen, 1983, p. 65).

While this might be a perspicuous definition of determinism, it is not likely understandable to participants. Hence, the use of the ‘complete cause’ operationalization of determinism trades off precision for interpretability. Note, however, that elements of the Universe A/B vignette retain some aspects of the entailment characterization of determinism. For example, the vignette includes the following gloss on determinism: “If everything in this universe was exactly the same up until John made his decision, then it had to happen that John would decide to have French Fries.” Under this description, it is true that, holding fixed the past and the laws of nature determines a unique state of affairs in the present. Thus, it is not completely inimical to the entailment characterization of determinism.

Further, the Supercomputer vignette describes determinism in terms of perfect prediction. Some have noted that determinism might not entail the possibility of perfect prediction in the actual world (Ismael, 2019). However, this objection misconstrues the force of the example. While determinism is not inherently related to prediction, and a universe where such a “perfect prediction” machine exists would be deterministic, even if there are deterministic universes where no such machine could exist. There is nothing about determinism that *rules out* the possibility of perfect prediction even though there is nothing about determinism that *requires* the possibility of perfect prediction. The fact that participants misunderstand at least one kind of determinism shows that researchers cannot simply assume that participants understand research materials in the way that researchers themselves do.

The simplest lesson to glean from our results is that the experimental study of free will suffers from a pervasive comprehension problem that problematizes extant findings. Only by controlling for epiphenomenalism, fatalism, intrusion and other potential sources of error can researchers keep their data about free will intuitions from being confounded. Whenever upwards of 80% of participants fail to understand key aspects of one’s experimental design, one must return to the drawing board. That is our prescription for experimental philosophers. We believe it is time to rethink how to best explore people’s intuitions about the relationship between free will and determinism. Insofar as participants consistently misunderstand the latter, their intuitions about the former will remain unreliable.\*

\*We consulted several of the leading philosophers who work on free will beliefs (from both the compatibilist and incompatibilist camps). We’d like to thank Shaun Nichols, Eddy Nahmias, Walter Sinnott-Armstrong, and Al Mele for their helpful suggestions when it comes to how our items for measuring epiphenomenalism, fatalism, and intrusion ought to be worded. We did our level best to address the concerns that people raised about earlier versions of the items we used to measure comprehension failures.

**References**

Andow, J., & Cova, F. (2016) Why compatibilist intuitions are not mistaken: A reply to Feltz and Millan, *Philosophical Psychology*, *29(4)*, 550–566. <https://doi.org/10.1080/09515089.2015.1082542>

Audi, R. (1991). Responsible Action and Virtuous Character. *Ethics,* *101*(2), 304-321. Retrieved October 26, 2020, from <http://www.jstor.org/stable/2381865>

Björnsson, G. (2014). Incompatibilism and “Bypassed Agency.” In A. Mele (Ed.), *Surrounding Free Will* (pp. 95–112). New York, NY: Oxford University Press.

Björnsson, G., & Pereboom, D. (2014). Free will skepticism and bypassing. In W. Sinnott-Armstrong (Ed.), *Moral Psychology: Free will and moral responsibility (Volume 4)* (pp. 27–36). Cambridge, MA: M.I.T. Press.

Burhmester, M., Kwang, T., and Gosling, S. (2011). Amazon’s Mechanical Turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science*, *6*, 3–5. [https://doi.org/10.1177/1745691610393980](https://doi.org/10.1177%2F1745691610393980)

Chihara, C.S., and Fodor, J. (1965) Operationalism and ordinary language: A critique of Wittgenstein. *American Philosophical Quarterly* 2:4, 281-295.

Cokely, E., & Feltz, A. (2010). Questioning the free will comprehension question. *Proceedings of the Annual Meeting of the Cognitive Science Society, 32*(32), 2440–2445.

Cover, J.A. and J. O’Leary-Hawthorne. (1996). Free agency and materialism. In J. Jordan & D. Howard-Snyder (Eds.), Faith, freedom and rationality (pp. 47-71). Lanham, MD: Roman and Littlefield.

Crone, D. & Levy, N. 2018. Are free will believers nicer people? Four studies suggest not. *Social Psychological and Personality Science*. [https://doi.org/10.1177/1948550618780732](https://doi.org/10.1177%2F1948550618780732)

De Brigard, F. Addis, D.R., Ford, J.H., Schacter, D.L., and Giovanello, K.S. 2013. Remembering what could have happened: Neural correlates of episodic counterfactual thinking. *Neuropsychologia* 51:12, 2401-2414.

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>

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). New York: Oxford University Press.

Dennett, D. (1984). I could not have done otherwise: So what? *Journal of Philosophy, 81*(10), 553–565.

Doris, J.M., & Knobe, J. (2010). Responsibility. In J.M. Doris (ed.), *The moral psychology handbook* (pp. 321–354). Oxford: Oxford University Press.

Ekstrom, L.W. (2002). Libertarianism and Frankfurt-style cases. In R. Kane (Ed.), *The Oxford handbook of free will* (pp. 309–322). New York: Oxford University Press.

Feltz, A., Cokely, E., & Nadelhoffer, T. (2009). Natural Compatibilism versus Natural Incompatibilism: Back to the Drawing Board. *Mind & Language, 24*(1), 1–23. <https://doi.org/10.1111/j.1468-0017.2008.01351.x>

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

Ismael, J. 2019. Determinism, Coounterpredictive Devices, and the Impossibility of Laplacean Intelligences. *The Monist* 1-2:4, 478-498.

Jackson, F. (1998). From metaphysics to ethics: A defense of conceptual analysis. New York: Oxford University Press.

Kane, R. (1999). Responsibility, luck, and chance: Reflections on free will and indeterminism. *Journal of Philosophy, 96,* 217-240.

Kane, R. (2005). *A Contemporary Introduction to Free Will.* New York: Oxford University Press.

Knobe, J. (2014). Free will and the scientific vision. In E. Machery & E. O’Neill (Eds.), *Current Controversies in Experimental Philosophy* (pp. 69–85). New York: Routledge.

Lycan, W. (2003). Free will and the burden of proof. In A. O’Hear (Ed.), Proceedings of the Royal Institute of Philosophy for 2001-02 (pp. 107–122). Cambridge, UK: Cambridge University Press.

Monroe, A. E., Dillon, K. D., & Malle, B. F. (2014). Bringing free will down to Earth: People’s psychological concept of free will and its role in moral judgment. *Consciousness and Cognition*, *27*, 100–108. <https://doi.org/10.1016/j.concog.2014.04.011>

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

Murray, S., Dykhuis, E., & Nadelhoffer, T. In preparation. Do people understand determinism? The comprehension problem for measuring free will beliefs.

Nadelhoffer, T., Rose, D., Buckwalter, W. & Nichols, S. (2020). Natural compatibilism, indeterminism, and intrusive metaphysics. *Cognitive Science*, *44:8*, e12873. <https://doi.org/10.1111/cogs.12873>

Nadelhoffer, T., Yin, S., & Graves, R. (2020). Folk intuitions and the conditional ability to do otherwise. *Philosophical Psychology* (online first)

<https://doi.org/10.1080/09515089.2020.1817884>

Nahmias, E. (2011). Intuitions about free will, determinism, and bypassing. In R. Kane (Ed.), *The*

*Oxford Handbook on Free Will* (2nd edition: pp. 555–576). New York: 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*, 561–584. <https://doi.org/10.1080/09515080500264180>

Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2006). Is incompatibilism intuitive? *Philosophy and Phenomenological Research*, *73*, 28–53. <https://doi.org/10.1111/j.1933-1592.2006.tb00603.x>

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). New York: Palgrave-Macmillan.

Nichols, S. (2006a). Folk intuitions on free will. *Journal of Cognition and Culture*, *6*, 57–86. <https://doi.org/10.1163/156853706776931385>

Nichols, S. (2006b) Free will and the folk: response to commentators. *Journal of Cognition and Culture*, *6*, 305–320. <https://doi.org/10.1163/156853706776931376>

Nichols, S. (2012). The indeterminist intuition. *The Monist*, *95*, 290–307. <https://doi.org/10.5840/monist201295216>

Nichols, S. (2015). Bound. Oxford: Oxford University Press.

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

Nowell-Smith, P.H. (1949). Free will and moral responsibility. Mind, 57: 45-65.

O’Connor, T. (2000). Persons and Causes: The Metaphysics of Free Will. New York: Oxford University Press.

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: 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. (2017). Neuroscientific prediction and the intrusion of intuitive metaphysics. *Cognitive Science*, *41*, 482–502. <https://doi.org/10.1111/cogs.12310>

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

Roskies, A., & Nichols, S. (2008). Bringing moral responsibility down to Earth. *The Journal of Philosophy*, *105*, 371–388. <https://doi.org/10.5840/jphil2008105737>

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.

<http://dx.doi.org/10.2139/ssrn.2160588>

Strawson, G. (1986). Freedom and Belief. Oxford: Oxford University Press.

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. (1983). *An Essay on Free Will*. Oxford: Oxford University Press.

Vargas, M. (2013). Building Better Beings. Oxford: Oxford University Press.

Vargas, M. (2017). Contested terms and philosophical debates. *Philosophical Studies* 174:10, 2499-2510.

Woolfolk, R. L., Doris, J. M., & Darley, J. M. (2006). Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition*, *100*, 283–301. <https://doi.org/10.1016/j.cognition.2005.05.002>

1. Other incompatibilists who appeal to commonsense thinking about free will include Ekstrom (2002); Kane (1999); O’Connor (2000); Pereboom (2001); and Strawson (1986). Other compatibilists who appeal to commonsense thinking about free will include Lycan (2003); Nowell-Smith (1949). [↑](#footnote-ref-1)
2. This paper presumes that folk intuitions sometimes provide relevant constraints on theory construction. This might be because folk intuitions are relevant for fixing the reference of theoretical terms (as in Vargas, 2017; see also Chihara & Fodor, 1965), or because congruence with folk intuitions might constitute a theoretical virtue that warrants (all else being equal) selecting a more congruent theory over a less congruent theory. Whatever the case may be, our research is situated within an ongoing debate about whether folk thinking about free will is best characterized as compatibilist or incompatibilist, which includes the assumption that such thinking is sometimes theoretically relevant. [↑](#footnote-ref-2)
3. Deery, Davis, & Carey (2015) found that participant responses to different scale items indicated context-sensitive commitments to both compatibilism and incompatibilism (see p. 791). This might appear to be thin evidence at best for natural compatibilism. However, compatibilism is the thesis, roughly, that free will and determinism are logically compatible. The results from Deery et al. suggest that many people do not think determinism *per se* precludes having free will, but only that determinism precludes having free will in certain circumstances. This implies that there are compatibilist elements in folk psychological thinking about free will, even if such thinking is not *uniformly* compatibilist. [↑](#footnote-ref-3)
4. For an early diagnosis of the problem, see Cokely and Feltz (2010). Their focus is on why so many people fail what we are calling ‘surface’ comprehension checks (see §2.1.1.). We are interested instead in how many people *understand* determinism. [↑](#footnote-ref-4)
5. To say that people falsely believe that determinism entails epiphenomenalism is not to say that people are disposed to state this entailment when prompted; rather, to say that people falsely believe that determinism entails epiphenomenalism is to say that responses elicited under experimental conditions are best explained by positing a false belief that determinism entails epiphenomenalism. [↑](#footnote-ref-5)
6. Feltz and Milan (2013) found that a sizable percentage of their participants believe that fatalism is *compatible* with free will and moral responsibility, which they take as evidence that many people believe in “free will no matter what” (see, also, Feltz, Cokely, & Nadelhoffer, 2009). [↑](#footnote-ref-6)
7. ‘□’ denotes the necessity operator, ‘→’ denotes entailment, and ‘&’ denotes the conjunction operator. ‘Po’ names some proposition that denotes the state of the world at time o, ‘L’ names some proposition that denotes a set of laws of nature, and ‘P’ names some proposition that denotes the state of the world at the present time. Informally, Po stands for “the past”, L stands for “the laws”, and P stands for “the present”, so that the above could be read roughly as noting the mistake of equating: “Necessarily, the past and the laws entail the present” with “The past is necessary, the laws are necessary, and together they entail the present”. Determinism does not rule out the possibility of alternative laws of nature or past sequences of events. [↑](#footnote-ref-7)
8. <https://osf.io/bvf4a/?view_only=59e488226fc34f39b214f799f9132fa4> [↑](#footnote-ref-8)
9. To determine sample size, we conducted a power analysis using G\*Power. For a linear regression with two predictors (vividness and either Epiphenomenalism, Fatalism, and intrusion predicting free will or moral responsibility) to detect a minimum effect size of interest (*d* = 0.15) at 95% power (error rate = 5%), we needed 107 participants. Given that only 1/3 of the sample would be administered each of the dependent variables of interest, we multiplied by a factor of 3 to get 321 participants. We over-recruited by 15% to account for exclusions. [↑](#footnote-ref-9)
10. 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, Kwang, & Gosling (2011); Paolacci, Chandler, & Ipeirotis (2010); Rand (2012). [↑](#footnote-ref-10)
11. This study was approved by the Institutional Review Board at the College of Charleston. [↑](#footnote-ref-11)
12. All stimuli and measures were presented on the same page for the study. This means that participants could always look back at the scenario to see the details. [↑](#footnote-ref-12)
13. A surface comprehension check can detect whether participants are able to correctly identify basic features of how the scenarios are worded but cannot detect whether they have a deeper understanding of the nature and implications of the key elements of the scenario. [↑](#footnote-ref-13)
14. As we noted above, 6% (19/325) failed this surface comprehension check. [↑](#footnote-ref-14)
15. Measures of vividness were included to explore the relationship between the quality of mental imagery and responses to items about the vignettes, but there were no pre-registered hypotheses about the relationship between vividness and judgments of free will. The 5-pt. Likert scale is a standard tool to measure vividness of mental imagery (see, e.g., De Brigard et al., 2013). Because analyses of vividness responses were merely exploratory, we do not discuss the results here. We mention them only to be fully transparent about all measures used in this research. [↑](#footnote-ref-15)
16. This type of ‘suspension of disbelief’ paradigm was originally used in Nahmias et al. (2005; 2006). [↑](#footnote-ref-16)
17. In their original study, Nichols and Knobe only asked about moral responsibility, which limited their ability to speak to the debate about natural incompatibilism, so we included statements about both free will and responsibility. [↑](#footnote-ref-17)
18. All six items were presented as a group and their order was randomized. [↑](#footnote-ref-18)
19. In designing these four items, we tried to word them in a way that addresses the main objection raised by Björnsson (2014) and Björnsson & Pereboom (2014) about earlier work on epiphenomenalism—namely, that the wording of the epiphenomenalism items used by Murray and Nahmias (2014) and Nahmias and Murray (2011) allows for an *incompatibilist* reading of their findings. Given the way we have worded our items (and given our results), we believe have sidestepped this possible strategy for explaining away our findings. [↑](#footnote-ref-19)
20. We included items that expressed both naïve and modally sophisticated forms of fatalism. [↑](#footnote-ref-20)
21. This information was collected in Study 1 and Study 2 to explore whether the responses from naïve participants significantly differ from experienced participants since past research suggests experience may make a difference (Crone & Levy, 2018). Because this element of our study was exploratory and we did not make any preregistered predictions, we do not discuss these findings here. The complete data set can be found on our OSF page. [↑](#footnote-ref-21)
22. Percentages for free will and responsibility judgments were calculated by taking all participants who registered 5-7 on the free will item or responsibility item as agreeing that people could have free will or be morally responsible in a deterministic universe, while participants who registered 1-3 were counted as disagreeing that people could have free will or be morally responsible in a deterministic universe. Ratings of 4 were not counted either way (hence, percentages do not add up to 100). We round percentages down (.1-.4) or up (.5-.9) accordingly throughout the paper. We dichotomized responses made on a continuous scale for purposes of comparing our results to those reported in Nichols and Knobe (2007). The data for free will and responsibility judgments was bimodally distributed, which suggested that the data could be represented ordinally. [↑](#footnote-ref-22)
23. Because our study design departed in some ways from Nichols and Knobe (2007), our results reflect a failed *conceptual* replication rather than a failed *direct* replication. A conceptual replication of some study attempts to find relationships between variables using slightly different methods than the original study to see whether empirical findings are robust to methodology. A direct replication of some study attempts to find relationships between variables using the exact same procedure as the original study. Thus, because we did not find a relationship between reading about a deterministic scenario and judgments about free will using slightly different measures, we failed to conceptually replicate Nichols and Knobe (2007). [↑](#footnote-ref-23)
24. It is surprising because most participants in that condition agreed with fatalism, so it is odd that this did not influence their intuitions about free will. [↑](#footnote-ref-24)
25. This study was approved by the Institutional Review Board at the College of Charleston. [↑](#footnote-ref-25)
26. The original Supercomputer scenario was set in the future in our own universe. We thought that by having participants instead imagine a generic universe where such a supercomputer existed, this would make it easier to ask how similar participants thought the Supercomputer scenario is to our own universe. [↑](#footnote-ref-26)
27. As we noted above, 5% (16/325) failed this surface comprehension check. [↑](#footnote-ref-27)
28. We would like to thank Al Mele for suggesting we include this item to control for the possibility that some people don’t find Jeremy blameworthy because they didn’t view the bank robbery negatively. [↑](#footnote-ref-28)
29. 96% of participants (291/302) judged that Jeremy’s robbing the bank was a bad thing. [↑](#footnote-ref-29)
30. One interesting difference between Studies 1 and 2 concerns the reliability of the error measures. While Intrusion items exhibited high reliability across studies, the reliability of Bypassing and Fatalism items improved markedly. In Study 1, the reliability of Fatalism items was driven down by Item 4. Dropping Item 4 raises the internal reliability of Fatalism items to .804 (Cronbach’s alpha). However, this difference in Fatalism items did not emerge in Study 2. We do not know why these differences emerged and whether they resulted from our sample or from some feature of the vignette. [↑](#footnote-ref-30)
31. Someone might view our method as stacking the deck against comprehension, where one mistake counts as misunderstanding. However, computing comprehension failures as a function of average responses to the four items does not significantly change results. We still observe high failure rates for epiphenomenalism (91% in S1; 43% in S2), fatalism (90% in S1; 67% in S2), and intrusion (57% in S1; 71% in S2) across both studies. The total failure rate using averages is 70% (409 participants). Hence, even on more conservative estimations of failure, we still observe high rates of misunderstanding. [↑](#footnote-ref-31)
32. Someone might still object to this interpretation of participant responses. After all, it might seem that we should utilize a principle that use fixes meaning, such that ‘determinism’ refers to whatever best explains the balance of participant responses in adequately powered, representative studies. On that view, we might wonder whether ‘determinism’ *just means* ‘fatalism’ or ‘epiphenomenalism’ or ‘bypassing’ given how participants behaved in our studies (more precisely, we might treat these terms as names of propositions, where ‘determinism’ and either of ‘fatalism’, ‘epiphenomenalism’, or ‘bypassing’ corefer to the same proposition). However, in Murray et al. (In preparation), we found that when participants are presented with measures from all three error scales, nearly 50% of participants made all three errors. Given that fatalism, epiphenomenalism, and bypassing are jointly inconsistent with each other, it seems more plausible to interpret participants as making erroneous judgments about determinism rather than fixing the reference of ‘determinism’ in a way that is consistent with the usage of the term. [↑](#footnote-ref-32)
33. It might seem the obvious response to invoke principles of reflective equilibrium. Empirical evidence (especially judgments taken to indicate folk conceptual commitments) must always be balanced against other evidence when drawing conclusions about concepts. However, this response misses the point. There is no work for reflective equilibrium to do here because there is nothing to bring into balance. Our results call into question the inferential value of extant evidence entirely. [↑](#footnote-ref-33)