

**Natural Compatibilism vs. Natural Incompatibilism:
Back to the Drawing Board**

Adam Feltz, Edward T. Cokely, and Thomas Nadelhoffer

Word Count (including abstract, references, and appendices): 10,962

(Please obtain a quotable copy from *Mind and Language*)

Abstract:

In the free will literature, some compatibilists and some incompatibilists claim that their views best capture ordinary intuitions concerning free will and moral responsibility. One goal of researchers working in the field of experimental philosophy has been to probe ordinary intuitions in a controlled and systematic way to help resolve these kinds of intuitional stalemates. We contribute to this debate by presenting new data about folk intuitions concerning freedom and responsibility that correct for some of the shortcomings of previous studies. These studies also illustrate some problems that pertain to all of the studies that have been run thus far.

In the free will literature, compatibilists and libertarians alike claim that their respective views best capture our ordinary intuitions concerning free will and moral responsibility. So what is the most useful way of understanding and addressing this sort of intuitional stalemate? One of the primary goals of researchers working in the field of experimental philosophy has been to probe intuitions in a controlled and systematic way in order to shed light on precisely these kinds of

Address for correspondence: Adam Feltz, 151 Dodd Hall, 641 University Way, P.O. Box 3061500, Tallahassee, FL 32306-1500.

Email: adf04@fsu.edu

debates.¹ In this paper, we contribute to this project by presenting results from a series of experiments that reveal flaws in all of the studies that have been run thus far. Along the way, we first provide an overview of some early attempts to probe folk intuitions that produced data that seemed to support Natural Compatibilism (NC)—i.e., the view that compatibilism² best settles with folk intuitions (Nahmias, Morris, Nadelhoffer, and Turner, 2005; 2006). Then we examine some recent follow-up studies that seem to support Natural Incompatibilism (NI)—i.e., the view that incompatibilism³ best settles with folk intuitions (Nichols and Knobe, forthcoming).⁴ Having examined the debate between these two competing camps, we then present the results of some new studies that improve upon the ones that have previously been used to defend NI. However, we will go on to suggest that all of the studies that have been proffered so far in support of NI—including our own—have fallen short of their original goal. Furthermore, we identify two problems that call into question the evidence supporting NC. If we are right, then it appears that experimental philosophers need to go back to the drawing board if they want to determine whether NC or NI is the correct view. We nevertheless hope to point the parties to the free will debate in the right direction by correcting for past experimental mistakes while at the same time hinting at possible directions for future research.

¹ For a thorough overview of recent work in experimental philosophy, see Nadelhoffer and Nahmias 2007.

² Compatibilism is the 'thesis that free action and moral responsibility are compatible with the truth of determinism' (Mele, 2006, p. 3). Determinism is the thesis that 'at any instant exactly one future is compatible with the state of the universe at that instant and the laws of nature' (Mele, 2006, p. 3).

³ Incompatibilism is the 'thesis that neither free action nor moral responsibility is compatible with the truth of determinism' (Mele, 2006, p. 3).

⁴ We want to emphasize that determining whether the folk have intuitions in accord with NI or NC is a descriptive project. We do not make the further claim that determining what the folk view is settles whether compatibilism or incompatibilism as philosophical views are correct.

1. Defending Natural Compatibilism

The traditional orthodoxy in the free will debate is that incompatibilism is the common view among the non-philosophical masses. Consider, for instance, the following remarks from Robert Kane:

In my experience, most ordinary persons start out as natural incompatibilists. They believe there is some kind of conflict between freedom and determinism; and the idea that freedom and responsibility might be compatible with determinism looks to them at first like a ‘quagmire of evasion’ (William James) or ‘a wretched subterfuge’ (Immanuel Kant). Ordinary persons have to be talked out of this natural incompatibilism by the clever arguments of philosophers. (1999, p. 217)

Similarly, Jan Cover and John Hawthorne claim that:

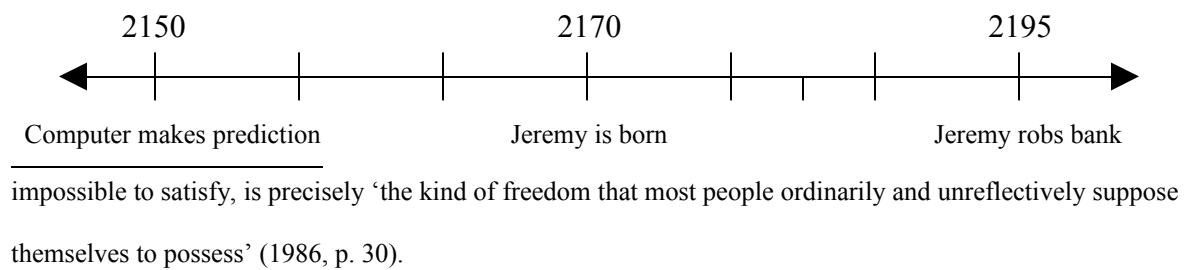
When ordinary people come to consciously recognize and understand that some action is contingent upon circumstances in an agent’s past that are beyond that agent’s control, they quickly lose a propensity to impute moral responsibility to the agent for that action. We can readily explain this fact by supposing 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)⁵

⁵ Other examples include Laura Ekstrom’s assertion that ‘we come to the table, nearly all of us, as pre-theoretic incompatibilists’ (2002, p. 310) so that ‘the compatibilist, then, needs a positive argument in favor of the compatibility thesis’ (2000, p. 57); and Galen Strawson’s claim that the libertarian conception of free will, though

But just because many (if not most) philosophers believe that NI is the default view, it does not follow that is. Luckily, by running studies that are designed to get at the salient folk intuitions in a controlled and systematic way, researchers can try to determine the truth of NI—a task that experimental philosophers have recently begun to undertake.

For instance, in a series of papers, Eddy Nahmias, Stephen Morris, Thomas Nadelhoffer, and Jason Turner (henceforth, the Nahmias group) discuss their attempts to investigate folk intuitions concerning free will, determinism, and moral responsibility (Nahmias *et al.*, 2005; 2006). In one study participants received the following vignette:

Imagine that in the next century we discover all the laws of nature, and we 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 such a supercomputer existed, and it looks at the state of the universe at a certain time on March 25th, 2150 A.D., twenty years before Jeremy Hall is born. The computer then deduces from this information and the laws of nature that Jeremy will definitely rob Fidelity Bank at 6:00 PM on January 26th, 2195. As always, the supercomputer’s prediction is correct; Jeremy robs Fidelity Bank at 6:00 PM on January 26th, 2195. (Nahmias *et al.*, 2005, p. 566)



impossible to satisfy, is precisely ‘the kind of freedom that most people ordinarily and unreflectively suppose themselves to possess’ (1986, p. 30).

While some participants were asked whether they believed Jeremy robbed the bank 'of his own free will', others were asked whether Jeremy was morally responsible for robbing the bank. The results indicated that a significant majority of participants (76%) who received the first question judged that Jeremy robs the bank of his own free will, and 83% who received the second question judged that Jeremy was blameworthy for robbing the bank.⁶ On the surface at least, these results suggest that people's pre-theoretical intuitions may be less incompatibilistic than philosophers have traditionally assumed.

At this point, one might object that in the Jeremy cases, the Nahmias group did not make the deterministic nature of the scenario salient enough to the participants. Perhaps participants were more focused on the fact that Jeremy's actions were predicted by the supercomputer than the fact that the prediction was made based on deterministic laws. To explore this possibility, they developed another scenario that presented determinism in a way that made it clear that the agents' behavior is sufficiently caused by factors beyond their control (i.e., genes and upbringing). In this follow-up study, participants were presented with the following vignette:

Imagine there is a world where the beliefs and values of every person are caused completely by the combination of one's genes and one's environment. For instance, one day in this world, two identical twins, named Fred and Barney, are born to a mother who puts them up for adoption. Fred is adopted by the Jerksens and Barney is adopted by the Kindersens. In Fred's case, his genes and his upbringing by the selfish Jerksen family have caused him to value money above all else and to believe it is OK to acquire money

⁶ The Nahmias group got similar results with vignettes that involved morally positive actions—e.g., saving a baby—and morally neutral cases—e.g., going jogging (2006, p.15). These results are all the more puzzling given the results of the studies run by Nichols and Knobe. But we will say more about that later in the paper.

however you can. In Barney's case, his (identical) genes and his upbringing by the kindly Kinderson family have caused him to value honesty above all else and to believe one should always respect others' property. Both Fred and Barney are intelligent individuals who are capable of deliberating about what they do. One day Fred and Barney each happen to find a wallet containing \$1000 and the identification of the owner (neither man knows the owner). Each man is sure there is nobody else around. After deliberation, Fred Jerkson, because of his beliefs and values, keeps the money. After deliberation, Barney Kinderson, because of his beliefs and values, returns the wallet to its owner. Given that, in this world, one's genes and environment completely cause one's beliefs and values, it is true that if Fred had been adopted by the Kindersons, he would have had the beliefs and values that would have caused him to return the wallet; and if Barney had been adopted by the Jerksens, he would have had the beliefs and values that would have caused him to keep the wallet. (Nahmias *et al.*, 2005, p. 570)

Despite providing participants with a robust description of agents whose actions were completely caused by their genes and environment, a significant majority of participants (76%) nevertheless judged both that Fred kept the wallet of his own free will and that Barney returned it of his own free will. This response pattern is consistent with the participants' responses in the aforementioned Jeremy studies, suggesting that this scenario probed similar intuitions about the relationship between determinism and free will.⁷

⁷ The Nahmias group also ran a third survey involving a roll-back scenario whereby the universe was re-created over and over such that given the same initial conditions and laws of nature, everything must happen the same way every time. Most participants judged that the agent acted of her own free will (66%), and most judged her to be morally responsible for her action (77%) (Nahmias *et al.*, 2006, p. 38).

When taken together, the results from the Nahmias group's early studies are not only *consistent* with the truth of NC, they also appear to be *inconsistent* with the claim that most people recognize that 'there is some kind of conflict between freedom and determinism' (Kane, 1999, p. 218)—i.e., they appear to be inconsistent with the truth of NI. Unfortunately, as is so often the case in philosophy, matters are more complicated than they at first appear. The results from some follow-up studies by Shaun Nichols and Joshua Knobe suggest that incompatibilism is more intuitive to non-philosophers than compatibilism after all. So, we should now turn our attention to their interesting research program.⁸

2. Defending Natural Incompatibilism

Following up on some of Nichols' early studies—which explored whether children believe in something akin to agent causation⁹ (Nichols, 2004)—Nichols and Knobe ran new studies in an effort to establish that NI rather than NC has more empirical support. On their view, most people are theoretically incompatibilist. However, Nichols and Knobe suggest that people can nevertheless be led to have compatibilist intuitions if they are presented with affectively charged scenarios. As they say, 'when people are confronted with a story about an agent who performs a

⁸ The first studies that appeared, on the surface, to support NI were run by Nichols (2004). But since these early studies focused on children, we are not going to examine them here. For even if children happen to have intuitions that comport with incompatibilism, it does not follow that the intuitions of adults are similarly incompatibilistic. Hence, these studies cannot help us adjudicate between NC and NI.

⁹ Agent causal accounts can be described as ones where 'an agent is in the strict and literal sense an originator, an initiator, an ultimate source of her directly free action; she is an uncaused cause of that behavior, and one whose causing of that behavior is not causally determined' (Clarke, 2003, p. 134).

morally bad behavior, this can trigger an immediate emotional response, and this emotional response can play a crucial role in their intuitions about whether the agent was morally responsible' (Nichols and Knobe, forthcoming). If affect is responsible for our compatibilist intuitions, then that should give defenders of NC pause for concern.

As Nichols and Knobe correctly point out, 'previous studies of people's moral responsibility intuitions all featured determinate agents and therefore were designed in a way that would tend to trigger affective reactions' (forthcoming). Hence, because all previous studies used determinately identified agents (e.g., Nahmias *et al.*, 2005; 2006; Woolfolk, Doris, and Darley, 2006), the participants' latent theoretical commitment to incompatibilism may have been distorted by the affect produced by the vignettes. On this view, if participants' compatibilist intuitions are driven by affect and not evidential features of the cases, then these intuitions may not give us any insight into their actual underlying theory concerning the relationship between free will, moral responsibility, and determinism.

In order to test their hypothesis, Nichols and Knobe (forthcoming) ran two different studies. The first presented participants with either abstract cases that were 'designed to trigger abstract, theoretical cognition' or concrete cases that were 'designed to elicit greater affective response' (Nichols and Knobe, forthcoming). Using a between-subject design, participants were first presented with descriptions of the following two universes:

Universe A: 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. (Nichols and Knobe, forthcoming).

Universe B: 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. (Nichols and Knobe, forthcoming)

Then one group of participants received the following concrete condition:

In Universe, A, a man named Bill has become attracted to his secretary, and he decides that the only way to be with her is to kill his wife and 3 children. He knows that it is impossible to escape from his house in the event of a fire. Before he leaves on a business trip, he sets up a device in his basement that burns down the house and kills his family. Is Bill fully morally responsible for kill his wife and children? (Nichols and Knobe, forthcoming)

In this high affect concrete condition, 72% of the participants judged that Bill was morally responsible for killing his wife. A separate group of participants were given the following low affect abstract condition 'In Universe A, is it possible for a person to be fully morally responsible

for their actions'? In this low affect condition, 86% percent of the participants judged that moral responsibility is not possible in Universe A. These results suggest that when presented with abstract scenarios most people have incompatibilist rather than compatibilist intuitions.

Nichols and Knobe tentatively think that the best explanation of their data is that an affective performance error drives participants' compatibilist intuitions in the concrete conditions. On their view, when participants are 'faced with truly egregious violations of moral norms...they experience a strong affective reaction which makes them unable to apply the theory correctly' (Nichols and Knobe, forthcoming). The bias generated by the high affect purportedly prevents participants from seeing that determinism rules out moral responsibility—something participants in the abstract condition do *not* overlook. Nichols and Knobe go on to offer some additional empirical support for their affective performance error model.

In a second study, participants received only concrete cases—except this time there were four different versions. Participants were first told that their case was set in either Universe A or Universe B, and they were then given one of the following conditions:

High Affect: As he has done many times in the past, Bill stalks and rapes a stranger. Is it possible that Bill is fully morally responsible for raping the stranger?

Low Affect: As he has done many times in the past, Mark arranges to cheat on his taxes. Is it possible that Mark is fully morally responsible for cheating on his taxes? (Nichols and Knobe, forthcoming)

Participants who were presented with cases in Universe B judged that the person in the scenario was morally responsible in both conditions (High Affect 95%, Low Affect 89%). Hence, it does not appear that affect influenced participants' ascriptions of responsibility in the indeterministic

conditions. However, affect did influence participants' responses in the deterministic conditions. In the high affect deterministic condition, 64% of the participants judged that Bill was morally responsible. In the low affect deterministic condition, only 23% of the participants claimed that Mark was responsible.¹⁰

When taken as a whole, the data collected by Nichols and Knobe suggest that while participants' ascriptions of moral responsibility may sometimes be compatibilist, their underlying theory concerning moral responsibility is incompatibilist. If this were correct, it would indeed provide strong evidence for the truth of NI. However, there are good reasons for thinking that perhaps the conclusions drawn by Nichols and Knobe may go beyond the scope of their actual data. To see why, we are now going to examine new data from studies based on Nichols and Knobe's experimental design—studies that put pressure on their affective performance error model.

3. Testing the Nichols and Knobe Hypothesis

Taking our cue from developments in other philosophical domains, we thought that perhaps there could be multiple folk concepts of moral responsibility at play rather than merely one. For example, in the action theory literature, both Fiery Cushman and Alfred Mele (forthcoming) and

¹⁰ It is worth pointing out that even if Nichols and Knobe are right that affect can help generate compatibilist intuitions in cases with high affect, their explanation does not explain why folk tend to give the same responses to low affect cases. After all, according to the studies from the Nahmias group., 79% of people responded that going jogging can be done freely in a deterministic universe (2006, p. 39). Because jogging appears not to be affectively charged, the affective performance error model seems unable to explain that result.

Shaun Nichols and Joseph Ulatowski (2007) have persuasively argued that there may be at least two folk concepts of intentional action (see also Feltz and Cokely, 2007). Similarly, Nichols, Jonathan Weinberg, and Stephen Stich (Nichols, Weinberg, and Stich, 2003; Weinberg, Nichols, and Stich, 2001) have provided evidence that there may be different folk epistemic concepts as well. Given that there appear to be multiple concepts in these other philosophically relevant domains, we thought that the same could be true of free will and moral responsibility. On this view, 'the' folk are neither monolithically compatibilist nor incompatibilist. Instead, there may be discrete groups of people who have different concepts of free will and moral responsibility and who therefore consistently express either compatibilist or incompatibilist intuitions.

In an effort to put the multiple concept view to the test, we borrowed Nichols and Knobe's vignettes and ran a study that relied on a within-subject design rather than the between-subject design that had been used in their original studies. We hypothesized that if participants were given both the high and low affect conditions, then groups of people would give matched responses to both conditions—which would indicate that discrete groups of people have stable intuitions. To test this view, we ran a study whereby participants received (a) descriptions of both the indeterministic and the deterministic universes, and (b) both high affect and low affect conditions.¹¹ In this study, 52 undergraduates at Florida State university volunteered to participate in the survey. The results turned out roughly as we had predicted. Twenty-five percent of the participants gave compatibilist answers to both questions, 67% gave incompatibilist answers to both questions, and only 8% gave mixed answers. As predicted, a large percentage of participants gave matched answers to both the high and low affect conditions. More importantly, most participants consistently gave either compatibilist or

¹¹ The vignettes were counterbalanced for order. An alpha level of 0.05 was chosen for all tests.

incompatibilist matched answers.¹² We think these results provide prima facie evidence that perhaps there is not one monolithic folk concept of moral responsibility, but rather a multiplicity of concepts.

If Nichols and Knobe's affective performance error model were right, then there should have been a clear order effect with respect to participants' responses to our follow-up study—i.e., there should have been more compatibilist friendly responses when the high affect condition was presented first than when the high affect was presented second. However, participants' responses were stable regardless of whether the high affect condition was presented first or second.¹³ But who is to say that participants didn't simply go back and revise their previous answers once they had read both prompts? We checked for this possibility and found a non-significant number of people changed their answers.¹⁴ Nevertheless, one explanation for why our participants' intuitions differed from the intuitions of the participants in the Nichols and Knobe studies is that our participants read *both* the cases *before* answering—a possibility that was not ruled out given our experimental design. Consequently, they may have realized that their intuitions in the two cases were inconsistent.

On this interpretation of the data, given that participants were cognizant of this intuitional conflict, they may have attempted to reconcile the inconsistency via something akin to reflective equilibrium.¹⁵ For instance, some participants may have recognized that there is a conflict

¹² $\chi^2(1, N=48) = 10.083, p < 0.01$

¹³ All F values were less than 1.

¹⁴ Participants were instructed that if they decided to change their answers, they should put an 'X' through their original answers and circle their new answer. Seven participants changed their answers.

¹⁵ There are two types of reflective equilibrium. Wide reflective equilibrium attempts to 'produce coherence in an ordered triple of sets of beliefs held by a person, namely, (a) a set of considered...judgments, (b) a set

between judging that a person who stabs his wife in Universe A is responsible whereas a person who cheats on his taxes is not. So, these participants may have concluded that because the 'world' is the same in all relevant respects in both cases, if one cannot be responsible for cheating on his taxes, then one cannot be responsible for stabbing his wife. Therefore, these participants would be inclined to give matched incompatibilist responses in their efforts to be consistent. Conversely, some participants may have judged that if someone is responsible for stabbing his wife in Universe A, then he is also responsible for cheating on one's taxes in Universe A. Hence, these participants would be inclined to give matched compatibilist responses. If participants read both scenarios, engaged in a process of reflective equilibrium, and then gave a reasoned judgment, then there is even more prima facie evidence that most people in this sample have stable incompatibilist intuitions whereas a minority have stable compatibilist intuitions.¹⁶

of...principles, and (c) a set of relevant background theories' (Daniels 1979, p. 258). Narrow reflective equilibrium attempts to find coherence between a set of considered judgments and a set of principles without consulting background theories (Daniels, 1979, p. 258). Given the disparity between the results of our studies and those of Nichols and Knobe, either type of reflective equilibrium best explains the difference between our results and theirs. The alternative is that the results reported by Nichols and Knobe were due to a sampling bias.

¹⁶ In an attempt to engage the participants in precisely this kind of reflective equilibrium, Nichols and Knobe performed a pilot experiment (forthcoming). They gave participants the abstract case, the high affect concrete case, and the results of their experiments. They explained that most people gave incompatibilist responses to the abstract scenario, most gave compatibilist responses compatibilist to the high affect concrete case, and that this pattern of responses is inconsistent. They asked participants to determine which view is correct—the compatibilist responses to the high affect case or the incompatibilist response to the abstract case. Nicholas and Knobe hypothesized that if participants had the chance to reflect on the inconsistency, they would come to some consensus. However, their prediction was not borne out by the results—even after reflection, 10 participants thought the incompatibilist response was correct and 9 thought the compatibilist response was correct (Nichols

Whatever the explanation, it is clear that when participants are given both the high and low affect conditions, not many people succumb to the affective performance error identified by Nichols and Knobe. After all, 67% of participants gave incompatibilist responses to both questions. Given these results, it appears on the surface that we have provided additional evidence for thinking that NI may be the correct view. After all, whether one uses a between-subject or a within-subject experimental design, the data suggest that people's intuitions are overwhelmingly incompatibilist—especially in cases involving low affect. Should we therefore conclude that NI is the default view? We do not believe that such a conclusion is warranted. In the following section we are going to examine some serious shortcomings of both our initial study and the studies run by Nichols and Knobe—shortcomings that threaten to undermine their evidentiary status with respect to the truth of NC and NI.

4. Getting from Semi-Compatibilism to Incompatibilism

Perhaps the most glaring problem with the studies discussed in the past two sections is that participants were only asked questions about moral responsibility and not free will.¹⁷ Consider,

and Knobe, forthcoming).

¹⁷ Recall incompatibilism states that moral responsibility and freedom are incompatible with determinism. In a footnote, Nichols and Knobe acknowledge that they are only concerned with whether people believe that moral responsibility and determinism are compatible and not with free will (forthcoming). However, they still go on to draw conclusions about incompatibilism—i.e., the view that free will and moral responsibility are incompatible with determinism—rather than semi-compatibilism—i.e., the view that moral responsibility is compatible with determinism even though free will is not. See Fischer and Ravizza 1998 for a defense of semi-compatibilism. Minimally, the fact that they use the term incompatibilism unnecessarily muddies the waters.

for instance, the questions Nichols and Knobe used in their high and low affect conditions. Participants were asked whether Bill and Mark could be fully morally responsible in Universe A and not whether they act of their own free will. Consequently, at best, the results of the studies by Nichols and Knobe shed light on what we are going to call for present purposes Natural Semi-Compatibilism (NSC)—i.e., the view that the majority of people believe that moral responsibility is compatible with determinism even if free will is not. The same can be said about our own follow-up study—which inherited this problem from the experimental design used by Nichols and Knobe. It is clear that in order to check whether people think NI—rather than NC or NSC—best captures folk intuitions, participants would need to be asked about *free will*.

In order to correct for this shortcoming, we ran a new study. Using Nichols and Knobe’s scenarios, we once again presented participants with descriptions of Universe A and Universe B. But this time they were asked specifically about free will rather than moral responsibility. In our effort to probe their intuitions concerning the relationship between free will and determinism, we relied on the following questions. Participants received both questions which were counterbalanced for order (see Appendix II for complete survey):

FW High: Bill lives in Universe A. As he has done many times in the past, Bill stalks and rapes a stranger. Does Bill rape the stranger of his own free will?

FW Low: Mark lives in Universe A. As he has done many times in the past, Mark arranges to cheat on his taxes. Does Mark cheat on his taxes of his own free will?

We hypothesized that participants would again show stable preferences in response to the high and low affect free will conditions as was observed in the high and low affect responsibility conditions.

In our first experiment, 65 undergraduates from low level philosophy courses at Florida State University volunteered to participate. The results from our revised study closely track the results of our earlier one. When participants were asked specifically about free will, 29% ($N = 19$) gave compatibilist matched responses, 62% ($N = 40$) gave incompatibilist matched responses, and 9% ($N = 6$) gave mixed responses. The difference between matched compatibilist and incompatibilist responses was statistically significant.¹⁸ Moreover, there was no significant general order effect present¹⁹, and order did not differentially affect either compatibilist or incompatibilist matched answers.²⁰ Finally, we tried to keep track of 'changed answers' by asking participants to put an 'X' through their original answers and circle their new answer. This allowed us to loosely measure whether people were reconsidering their answers. In this study, only three participants changed their answers. Whereas two participants changed their answers in the FW low affect condition from 'yes' to 'no' when the high affect condition was presented *second*, the other participant changed the answer from 'yes' to 'no' when the high affect condition was presented *first*.²¹

We then ran a second study in order to replicate our results with increased power. This time, 110 undergraduates at Florida State University volunteered to take part in the experiment.

¹⁸ $\chi^2(1, N=59) = 7.475, p < 0.01$

¹⁹ $\chi^2(1, N=59) = 0.857, p > 0.05$

²⁰ When FW High is presented first, 18 participants gave incompatibilist responses whereas 22 gave incompatibilist responses when FW High is presented second, $\chi^2(1, N=40) = 0.4, p > 0.05$. When FW High is presented first, 11 gave compatibilist answers while 8 gave compatibilist answers when presented FW High is presented second, $\chi^2(1, N=19) = 0.474, p > 0.05$.

²¹ Here again, it's possible that the participants read both vignettes before answers. But if so, we once again think this strengthens rather than weakens our overall interpretation of the data.

They received the same materials that the participants received in the first study. The results of this second study were nearly identical to those of the first—namely, 29% ($N=32$) gave compatibilist matched responses, 63% ($N=69$) gave incompatibilist matched responses, and 8% ($N=9$) gave mixed answers. Once again there was a significant difference between those who gave compatibilist and those who gave incompatibilist matched responses.²² Moreover, there was no general order effect²³ and no order effect with respect to the matched answers.²⁴ By our lights, these results put additional pressure on Nichols and Knobe's affective performance error explanation of compatibilist intuitions.

To illustrate, consider that, according to Nichols and Knobe, affect can cause the same *individual* to express compatibilist intuitions in one scenario and incompatibilist intuitions in another. As Nichols has recently suggested:

Joshua Knobe and I maintain that...the folk can be led to report compatibilist intuitions and they also be led to report incompatibilist intuitions. Our suspicion was that in the absence of emotional triggers, people would respond as incompatibilists. But we also thought that by making the scenario emotionally salient, people would be more likely to respond as compatibilists. (Nichols, 2006a, p. 75)

On this view, 'under some conditions, people's intuitions tend to compatibilism; under other

²² $\chi^2(1, N=101)= 13.554, p<0.01$.

²³ $\chi^2(1, N=101)= 2.784, p>0.05$.

²⁴ 33 gave incompatibilist answers when FW High was presented first, and 36 gave incompatibilist when FW high was presented second $\chi^2(1, N=67)= 0.130, p>0.05$. 21 gave compatibilist answers when FW High was presented first, 11 gave compatibilist answers when FW High was presented second $\chi^2(1, N=33)= 3.667, p>0.05$. It should be noted that there was a near significant effect for compatibilist matched answers.

conditions they tend to incompatibilism' (Nichols, 2006a, p. 75). In this respect, Nichols and Knobe are essentially suggesting that in morally laden situations, people's intuitions may 'flip-flop' from incompatibilism to compatibilism.

In contrast, our studies suggest that only a small number of individuals have mixed intuitions. In each experiment, whereas only roughly 8% of the participants gave non-matched answers, the overwhelming majority gave matched answers. Hence, it appears that there are different *groups* of people who express consistent compatibilist or incompatibilist intuitions about these cases. This finding suggests that only a small number of individuals may fall prey to the affective performance error using the current design and Nichols and Knobe's materials. Given that the majority of participants gave matched answers to both the high and low affect conditions, it does not appear that affect was biasing their judgments. Our claim that groups of people express stable incompatibilist or compatibilist intuitions is bolstered by the fact that we did not find (a) a general order effect, (b) a matched answer order effect, or (c) a 'changed' answer order effect that would have indicated that participants corrected for what they perceived to be an 'error'.

If Nichols and Knobe's affective performance error model were correct, then the high affect conditions should have biased participants' intuitions in the compatibilist direction. However, our studies do not support this conclusion. On the one hand, when we used a within-subject experimental design to probe folk intuitions about moral responsibility, we found that most people gave incompatibilist answers to moral responsibility questions even in the high affect cases. On the other hand, when we used a similar experimental design to probe folk intuitions concerning free will, we found that when the high affect condition was presented first,

59% ($N=51$) gave matched incompatibilist responses, 37% ($N=32$) gave matched compatibilist responses, and only 5% gave mixed answers ($N=4$). When the low affect condition was presented first, 66% ($N=58$) gave matched incompatibilist-friendly responses, 22% ($N=19$) gave matched compatibilist-friendly responses, and only 13% ($N=11$) gave mixed responses. The results from both of our studies are statistically significant.²⁵ Unlike Nichols and Knobe, we did not find that high affect leads most people to have compatibilist intuitions. At least in our studies, most participants expressed incompatibilist intuitions about free will even in high affect cases. So, as it stands, while the majority of people appear to have consistently incompatibilist intuitions, a sizeable minority have consistently compatibilist intuitions.

On the surface, these results not only cause problems for Nichols and Knobe's affective performance error model, they also seem to provide additional evidence that NI accurately captures folk intuitions about free will and moral responsibility after all. However, there are reasons to suspect that the NI interpretation is premature. That is, all of the studies built upon Nichols and Knobe's experimental design may share a second serious flaw—one that threatens to further undermine the ability of these studies to shed light on NI. To see why, let's return once again to the wording of the original vignettes used by Nichols and Knobe.

5. Determinism, Fatalism, and Could Have Done Otherwise

As we saw earlier, Nichols and Knobe gave participants descriptions of Universes A and B and

²⁵ In order to increase our power to detect possible effects, the data sets of the two experiments were combined for additional analyses. FW High, $\chi^2(1, N=83)= 4.349, p<0.05$. FW Low, $\chi^2(1, N=77)= 19.753, p<0.01$. There is also no order effect present in this larger data set, $\chi^2(1, N=160)= 3.543, p>0.05$.

then told them the following:

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.

Participants were then asked 'which of these universes do you think is most like ours'? Quite predictably, the vast majority of participants answered that Universe B—i.e., the indeterministic universe—is most similar to our own. But by our lights, this is a misleading way of framing the issue for several reasons. First and foremost, the 'has to happen' vs. 'does not have to happen' language makes it sound as if in a deterministic universe everything that happens necessarily happens. As Turner and Nahmias have recently pointed out, 'it is important, of course, to distinguish the trivial claim that, in deterministic universes, nothing could have been different if the past and laws were not different, from the downright false claim that, in deterministic universes, everything had to happen as it did' (Turner and Nahmias 2006, p. 606).²⁶

Failing to make this important modal distinction is tantamount to conflating determinism and fatalism. After all, determinism merely entails that $[(Po \ \& \ L) \rightarrow P]$ —i.e., necessarily, given the actual previous state of the universe (Po) and the actual laws of nature (L), there is only one possible outcome (P). Fatalism, on the other hand, is the claim that $\square(Po, L, \text{ and } P)$ —i.e., it is impossible that the past, the laws, and the present state of affairs could have been otherwise. In light of this worry, Nahmias has recently suggested that:

The problem is that determinism should not be described in a way that suggests that actual events, including human choices, could not happen in any other way. This is

²⁶ See also Nahmias *et al.*, 2006.

because determinism does not entail that nothing could happen otherwise—that all actual events are necessary. The fact that an event X is completely caused (or determined) by prior events does not entail that X has to happen (necessarily happens). (Nahmias 2006, p.222)

Minimally, it is at least a live possibility that the 'had to happen/did not have to happen' language used by both Nichols and Knobe and by us may have led participants to conflate determinism and fatalism in precisely the way Nahmias suggests.²⁷

This possibility is especially worrisome given that virtually no one—not even the

²⁷ There is evidence that people are in fact sensitive to the wording of the scenarios that imply fatalism. Nahmias, Coates, and Kvaran (2007) conducted studies that described determinism in terms of complete causation yet did not use the 'has to happen' language. These scenarios 'described deterministic worlds in which agents' decisions are (1) completely caused by prior events, and (2) those prior events were completely caused by earlier events going back to events before the agent was born, such that (3) the prior events will definitely cause the later events' (Nahmias et al., 2007, p. 222). One of their targets was to determine if framing the abstract scenarios in psychologically non-reductionist terms (e.g., thoughts, desires, plans) would elicit different intuitions about one being free and responsible from scenarios framed in psychologically reductionist terms (e.g., chemical reactions, neural processes). They did indeed find this effect—people given the reductionist scenario thought that one was less free and responsible than one in the non-reductionist scenario. However, more important for the present purposes, Nahmias et. al. tested to see if affect played a role in intuitions about concrete cases. Affect did play some role. Agents who did something good were judged as less free and responsible than an agent who did something bad. But still, about 60 % participants respond that the person in the scenario has free will, and between 63-81% of participants thought the person was morally responsible regardless of whether they were framed in reductionist or non-reductionist terms (Nahmias et al., 2007, p. 227). This suggests that simply leaving out the phrase that one's action 'has to happen' as it does led to people having predominately compatibilist friendly intuitions.

compatibilist—thinks that we can be free and morally responsible in a fatalistic universe.²⁸ Hence, we cannot conclude that just because participants judge that agents in Universe A are not free and responsible that this is because the participants are incompatibilists—it may just mean they are *non-fatalists*. To put it differently, at best when the results of Nichols and Knobe’s original studies (which asked only about moral responsibility) are coupled with the results from our follow-up studies (which asked about free will) it appears that all that has been shown is that Natural Fatalistic Incompatibilism (NFI) accurately captures most participants’ intuitions concerning the relationship between fatalism, free will, and moral responsibility. But unfortunately, none of these data speak to the truth of either NC or NI.²⁹

The second main problem with the 'had to happen/did not have to happen' language is that it is too ambiguous with respect to what it means to say that an agent could have done otherwise. According to an unconditional analysis of 'could have done otherwise', to say that an agent P could have done otherwise at t^1 is to say that P could have done otherwise even if all of the antecedent external environmental conditions and internal psychological conditions had remained constant. Compatibilists, on the other hand, typically offer a conditional analysis of 'could have done otherwise' whereby to say that an agent P could have done otherwise at t^1 is just to say that (a) some of the antecedent conditions could have been different, and (b) had they been different, P could (and perhaps would) have done otherwise.³⁰ The key issue in this debate is

²⁸ Calvinists are one exception to the rule.

²⁹ We want to reemphasize that compatibilist friendly responses to Nichols and Knobe's scenarios do not necessarily reflect an error. For example, some people may genuinely believe that one is free and responsible even in a fatalistic universe—e.g., Calvinists. It is not a mistake these people make, but rather they have a belief that most compatibilists do not share.

³⁰ For a similar discussion of this issue as it pertains to Nichols earlier experimental work, see Turner and Nahmias

whether the antecedent conditions or laws could have been otherwise. Libertarians think that determinism forecloses this possibility whereas compatibilists do not. Attempting to resolve this thorny issue would take us too far a field for present purposes. Minimally, given that we do not know which analysis of 'could have done otherwise' the participants are relying on when making their judgments in the Nichols and Knobe-style cases, we cannot know whether their answers support NI rather than NC.

Indeed, this ambiguity may shed light on why so many participants in the studies run by the Nahmias group had compatibilist intuitions whereas so many participants in the follow-up studies by both Nichols and Knobe and by us had incompatibilist intuitions. In the former studies, participants were merely told that a computer predicted that Jeremy would definitely rob the bank—which leaves open the possibility that Jeremy could have done otherwise if either the laws of the universe or the state of the world had been slightly different. In the latter studies, on the other hand, the vignettes suggest that things could not have been otherwise. If read in this way, then *both* compatibilists and libertarians should give what Nichols and Knobe take to be the incompatibilist answer. Indeed, the surprising thing is that so many people had *compatibilist* intuitions! After all, how could we be free in a world in which none of our desires, beliefs, and decisions could have been otherwise?

Perhaps what these results show is that the belief in free will and moral responsibility is so deeply entrenched in the minds of some individuals that nothing—not even the apparent truth of fatalism—could dislodge their view. Our suspicion is that at least some of the participants who appear to give compatibilist answers in the studies that have been run so far have a world view where God has given them an immaterial soul that allows us to stand above the causal fray

(2006).

—a soul that is free and responsible *no matter what*. On this view, we can *always* do otherwise than we actually do. If this were correct, it would explain why so many participants often fail the manipulation checks in these kinds of studies.³¹

Given that some participants likely have a world view that entails they are 'free will not matter what', they may be unable to do the conditional reasoning necessary for imagining that their world operates deterministically (or fatalistically).³² Insofar as they are convinced that they possess an immaterial soul and the kind of robust free will that is required for moral responsibility, they may simply be unable to get their minds around the possibility of a deterministic world since the existence of such a world would undermine their entire world-view. So, even if they may be perfectly willing to say that creatures a lot like humans are not free and responsible in a deterministic world, it is only because these creatures—unlike humans—*do not have a soul*. Perhaps if the vignettes specified that God explicitly gave these creatures free will, then the 'free will not matter what' participants would no longer think that determinism poses a threat. The worry here is that a deeply entrenched world view where God has given us free will may prevent some participants from imagining a deterministic world.

But if some participants are unable to do the conditional reasoning necessary for comprehending the nature of determinism, then why shouldn't we simply discard their answers? After all, that's what manipulation checks are for in the first place—*viz.*, to identify people who do not understand the thrust of the study. The main reason not to simply discard these

³¹ This is a problem that we believe has not received the attention it deserves. But addressing here in detail would take us too far afield.

³² Of course, people who believe in "free will no matter what" may not think that people who are *manipulated* to act as they do are free. That is, there are boundary conditions on their notions of free will.

participants' answers is that in doing so, we may be discarding the answers of a sizeable number of people whose belief in free will is so strong they simply cannot imagine living in a deterministic world. To live in such a world, on this view, is tantamount to living in a world without God. After all, if their world view entails they could always do otherwise, to ask them to imagine that they *could not* do otherwise given the past and the laws of nature would be to ask them to deny the truth of their deeply held world view. If such people exist—and it is clearly reasonable to assume that they do—by simply ignoring their answers, we may be thereby artificially inflating the rate of compatibilism among the folk.

More importantly, the possibility of people with a world view that entails we are 'free will no matter what' creates problems for *all* of the experimental designs that have been used so far given that both compatibilists and those who believe in God-given free will be inclined to give the same answers to many of the vignettes. Whereas participants with compatibilist intuitions will say that an agent can be free and responsible even if the universe is determined, those participants who believe we are free and responsible no matter what will also say that the agent is free and responsible—but only because they are theoretically blinded to the determinism at play in the scenario. Some of these latter participants may get the manipulation correct—and hence, be counted as compatibilists—while some will miss the manipulation check and have their answers discarded altogether. Either way, we could end up with some incompatibilists masquerading as compatibilists while others end up having their answers left off the ledger. But this is a hypothesis that can only be tested with future research. Regardless of whether some people think we have free will no matter what, the key question for present purposes is whether most people have a conditional or an unconditional notion of 'could have done otherwise'.

This is an issue that Nichols has tried to address in a follow-up study. First, he cites Kane (1996) in support of his claim that conditional analyses are 'widely rejected in the contemporary free will literature' (Nichols 2006a, p. 306). Of course, whether a conditional analysis of 'could have done otherwise' is 'widely rejected' by contemporary philosophers is itself an empirical question calling for more experimental spadework than has been provided by either Kane or Nichols. But even if we were to grant for the sake of argument that both Kane and Nichols are right in thinking that most philosophers have and should reject the conditional analysis, this would nevertheless not shed any light on the present debate. What matters for the problem at hand is not whether *philosophers* find the conditional analysis plausible but whether *laypersons* are relying on such an analysis when they read the Nichols and Knobe-style vignettes.

This is an important issue that Nichols is keenly aware of, so rather than trying argument based on his own intuitions (or the intuitions of philosophers more generally), he ran a pilot study to probe folk intuitions concerning 'could have done otherwise'. In this study, participants were given the following:

On 4/13/2005, Bill filled out his tax form. At precisely 10:30AM, he decides to lie about his income. But of course he didn't have to make this decision. Bill could have decided to be honest. (Nichols, 2006b, p. 306)

Using a between-subject experimental design, two groups of participants were then asked whether one of the following two sentences 'sounded right or wrong' (on a scale of 3 to -3):

1. Bill could have decided to be honest at 10:30, 4/13/2005, but only if some things had been different before the moment of his decisions.
2. Bill could have decided to be honest at 10:30, 4/13/2005, even if nothing had been

different before the moment of his decision. (Nichols, 2006b, p. 306)

It turns out that the participants who were presented with the sentence modeled on the unconditional analysis gave a higher rating than those who were presented with the sentence modeled on the conditional analysis. And while noting that this pilot study is not decisive, Nichols takes these results to suggest that the worry about the conditional analysis raised by Nahmias and others may be over-inflated. However, we are not convinced this particular experiment sheds as much light as Nichols assumes given that the language used in these two questions is entirely too vague. After all, it is unclear whether the 'some things' that could have been different include the mental states of Bill at the time. A much better question would have been something along the following lines:

- A. Bill could have decided to be honest at 10:30, 4/113/2005, but only if either something in the world or some of his beliefs and desires had been different before the moment of his decision.
- B. Bill could have decided to be honest at 10:30, 4/113/2005, even if nothing had been different—not even his beliefs, desires, deliberations, etc.—before the moment of his decision.

Our worry with Nichols' original wording is that it does not make it clear enough that the 'something' that needed to be different for Bill to make a different decision includes his mental states.

That this worry is legitimate seems to be supported by an early study run by the Nahmias group that attempted to similarly probe folk intuitions concerning what it means to say that an agent could have done otherwise. In this earlier study, participants were presented with the

following vignette:

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). (Nahmias *et al.*, 2004, p. 174)

Participants were then asked the following question:

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. (Nahmias *et al.*, 2004, p. 174)

In response to these options, 62% of the participants chose the 'compatibilist response' (b), 35% offered the 'libertarian response' (a), and 3% answered 'neither' (c).

These results suggest that when we don't introduce any potentially biasing moral issues into the cases, and we also make it clear that the 'something' that could have been different includes the mental states of the agent, participants have judgments that accord with the conditional analysis after all. Of course, more research is needed to settle the issue in a satisfactory manner. As it stands, we minimally think that it is an open question whether participants rely on a conditional rather than an unconditional analysis of 'could have done

otherwise' when making judgments concerning the compatibility of free will and moral responsibility with determinism. Hence, we believe that it is still an open question whether NC or NI is the correct view with respect to folk intuitions concerning free will and moral responsibility.

6. Conclusion

In this paper, we set out to contribute to the on-going debate about Natural Compatibilism and Natural Incompatibilism by providing additional evidence for the latter position. Unfortunately, things did not turn out as we had originally planned. Rather than showing that most people have incompatibilist intuitions, we ended up identifying several methodological shortcomings with the studies that have been run thus far by experimental philosophers. First, some of these studies have not asked participants about both free will and moral responsibility and therefore cannot help us adjudicate between NC and NI. Second, other studies have done a poor job of making sure participants are reading the vignettes deterministically rather than fatalistically. Third, it is still an open question whether participants rely on a conditional analysis of “could have done otherwise” or an unconditional analysis. And finally, there is a real worry that a sizeable percentage of participants—who may have a “free will no matter what” mentality—are distorting our estimations concerning how many people have compatibilist rather than incompatibilist intuitions. In light of these problems, we believe that experimental philosophers need to go back to the drawing board if they want to resolve the debate between those who support NC and those who support NI.

Adam Feltz

Department of Philosophy

Florida State University

Edward Cokely

Center for Adaptive Behavior and Cognition

Max Planck Institute for Human Development

Thomas Nadelhoffer

Department of Philosophy

Dickinson College

Appendix I

Please read the following scenarios and answers the questions (1)-(9). If you decide to change an answer, please put an X over your initial answer and circle your new answer.

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 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.

1. Which of these universes do you think is most like ours? (circle one)

Universe A Universe B

2. Mark lives in Universe A. As he has done many times in the past, Mark arranges to cheat on his taxes. Is it possible that Mark is fully responsible for cheating on his taxes?

Yes No

3. How strongly do you feel about your answer to (2)?

1	2	3	4	5	6	7
Not Strongly At All			Moderately Strongly			Very Strongly

4. Bill lives in Universe A. As he has done many times in the past, Bill stalks and rapes a

stranger. Is it possible that Bill is fully morally responsible for raping the stranger?

Yes No

5. How strongly do you feel about your answer to (4)?

1	2	3	4	5	6	7
Not Strongly At All			Moderately Strongly			Very Strongly

6. John lives in Universe B. As he has done many times in the past, John arranges to cheat on his taxes. Is it possible that John is fully responsible for cheating on his taxes?

Yes No

7. How strongly do you feel about your answer to (6)?

1	2	3	4	5	6	7
Not Strongly At All			Moderately Strongly			Very Strongly

8. Richard lives in Universe B. As he has done many times in the past, Richard stalks and rapes a stranger. Is it possible that Richard is fully morally responsible for raping the stranger?

Yes No

9. How strongly do you feel about your answer to (8)?

1	2	3	4	5	6	7
Not Strongly At All			Moderately Strongly			Very Strongly

Appendix II

Thank you for participating. Please read the following scenarios and answers the questions. If

you decide to change an answer, please put an X over your initial answer and circle your new answer.

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 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.

1. Which of these universes do you think is most like ours? (circle one)

Universe A

Universe B

2. Bill lives in Universe A. As he has done many times in the past, Bill stalks and rapes a stranger.

Does Bill rape the stranger of his own free will?

Yes

No

3. How strongly do you feel about your answer to (2)?

1
Not Strongly
At All

2

3

4
Moderately Strongly

5

6

7
Very
Strongly

4. Mark lives in Universe A. As he has done many times in the past, Mark arranges to cheat on his taxes. Does Mark cheat on his taxes of his own free will?

Yes

No

5. How strongly do you feel about your answer to (4)?

1
Not Strongly
At All

2

3

4
Moderately Strongly

5

6

7
Very
Strongly

References

- Clarke, R. 2003: *Libertarian Accounts of Free Will*. New York: Oxford University Press.
- Cover, J. A., and O'Leary-Hawthorne, J. 1996: Free agency and materialism. In J. Jordan and D. Howard-Snyder (eds.), *Faith, Freedom, and Rationality*. Lanham: Roman and Littlefield.
- Cushman, F., and Mele, A. forthcoming: Intentional action: Two-and-a-half folk concepts? In J. Knobe and S. Nichols (eds.), *Experimental Philosophy*. Oxford: Oxford University Press.
- Daniels, N. 1979: Wide reflective equilibrium and theory acceptance in ethics. *Journal of Philosophy*, 76, 256-282.
- Ekstrom, L. 2000: *Free Will*. Boulder, CO: Westview Press.
- Ekstrom, L. 2002: Libertarianism and Frankfurt-style cases. In R. Kane (ed.), *The Oxford Handbook of Free Will*. New York: Oxford University Press.
- Feltz, A., and Cokely, E.T.: An anomaly in intentional action ascriptions: More evidence of folk diversity. In D. S. McNamara and J. G. Trafton (eds.), *Proceedings of the 29th Annual Cognitive Science Society*. Austin, TX: Cognitive Science Society.
- Fischer, J. M., and Ravizza, M. 1998: *Responsibility and Control*. New York: Cambridge University Press.
- Goldman, A. 1986: *Epistemology and Cognition*. Cambridge: Harvard University Press.
- Kane, R. 1999: Responsibility, luck, and chance: Reflections on free will and indeterminism. *Journal of Philosophy*, 96, 217-240.
- Kane, R. 1996: *The Significance of Free Will*. New York: Oxford University Press.
- Knobe, J. 2003: Intentional action and side-effects in ordinary language. *Analysis*, 63, 190-194.

- Mele, A. 2006: *Free Will and Luck*. New York: Oxford University Press.
- Nadelhoffer, T., and Nahmias, E. 2007: The past and future of experimental philosophy. *Philosophical Explorations*, 10, 123-149.
- Nahmias, E. 2006: Folk fears about freedom and responsibility: Determinism vs. reductionism. *Journal of Cognition and Culture*, 6, 215-237.
- Nahmias, E., Coates, D., and Kvaran, T. 2007: Free will, moral responsibility, and mechanism: Experiments on folk intuitions. *Midwest Studies in Philosophy*, 31, 214-242.
- Nahmias, E., Morris, S., Nadelhoffer, T., and Turner, J. 2004: The phenomenology of free will. *The Journal of Consciousness Studies*, 11, 162-179.
- Nahmias, E., Morris, S., Nadelhoffer, T., and Turner, J. 2005: Surveying freedom: Folk intuitions about free will and moral responsibility. *Philosophical Psychology*, 18, 561-584.
- Nahmias, E., Morris, S., Nadelhoffer, T., and Turner, J. 2006: Is incompatibilism intuitive? *Philosophy and Phenomenological Research*, 73, 28-53.
- Nichols, S. 2004: The folk psychology of free will: Fits and starts. *Mind and Language*, 19, 473-502.
- Nichols, S. 2006a: Folk intuitions on free will. *Journal of Cognition and Culture*, 6, 57-86.
- Nichols, S. 2006b: Free will and the folk: Response to commentators. *Journal of Cognition and Culture*, 6, 305-320.
- Nichols, S., and Knobe, J. forthcoming: Moral responsibility and determinism: The Cognitive science of folk intuition. *Nous*.
- Nichols, S., Stich, S., and Weinberg, J. 2003: Metaskepticism: Meditations in ethno-

- epistemology. In S. Luper (ed.), *The Sceptics: Contemporary Essays*. Burlington: Ashgate.
- Nichols, S., and Ulatowski, J. 2007: Intuitions and individual differences: The Knobe effect revisited. *Mind and Language*, 22, 346-365.
- Strawson, G. 1986: *Freedom and Belief*. Oxford: Oxford University Press.
- Turner, J., and Nahmias, E. 2006: Are the folk agent-causationists? *Mind and Language*, 21, 597-609.
- Weinberg, J., Nichols, S., and Stich, S. 2001: Normativity and epistemic intuitions. *Philosophical Topics*, 29, 429-460.
- Woolfolk, R., Doris, J., and Darley, J. 2006: Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition*, 100, 283-301.