BRINGING MORAL RESPONSIBILITY DOWN TO EARTH*

Thought experiments have played a central role in philosophical methodology, largely as a means of elucidating the nature of our concepts and the implications of our theories.¹ Particular attention is given to widely shared “folk” intuitions—the basic untutored intuitions that the layperson has about philosophical questions.² Folk intuition is meant to underlie our core metaphysical concepts, and philosophical analysis is meant to explicate or sometimes refine these naïve concepts. Consistency with the deliverances of folk intuitions is a sign that the philosopher is making contact with his object of interest. In order to explore folk concepts, people are often asked to provide their intuitions about a state of affairs in some alternate universe or possible world, one that differs in particular, precise ways from the way things are in the actual world. Here we provide evidence that people’s intuitions about moral responsibility sometimes diverge across worlds even when the facts about these worlds are the same. Which world one considers actual affects at least some philosophical judgments, suggesting that it is not just possible worlds to which our intuitions are tied. We will present several possible explanations for the asymmetry we have identified, and we will consider some implications for philosophical intuition.

It has frequently been claimed that the folk are incompatibilists about freedom and moral responsibility—that they believe that freedom is not possible in a deterministic universe, and that if you are not free, you are not morally responsible.³ Thus Robert Kane claims, “In my experience, most ordinary persons start out as natural incompatibilists.”⁴ And Derk Pereboom writes, “Beginning students typically

---

*The authors contributed equally to this work. We received helpful comments on this work from a number of people. We would like especially to thank David Braddon-Mitchell, John Fischer, Richard Holton, Joshua Knobe, Uriah Kriegel, Derk Pereboom, Walter Sinnott-Armstrong, Saul Smilansky, and Roy Sorensen.


³ As John Fischer has emphasized in “Recent Work on Moral Responsibility” Ethics, cx (1999): 93–139, it is possible that determinism is consistent with responsibility but not with free will. Thus, it is often important to distinguish between moral-responsibility compatibilism and free will compatibilism, and we will keep the views distinct here.

recoil at the compatibilist response to the problem of moral responsibility. Of course, other philosophers have suggested that the common view is actually compatibilist.

The nature of our intuitions about free will and moral responsibility is not, however, purely a matter of a priori debate. What the folk think is an empirical question, and one which can be addressed by what is coming to be known as “experimental philosophy.” Recently, experimental philosophical approaches have yielded conflicting results as to whether the folk are, in general, compatibilists or incompatibilists about moral responsibility. Eddy Nahmias and colleagues had subjects assume that determinism is true, and then judge whether an agent is blameworthy under those circumstances. They found that subjects tended to say that the agent was blameworthy. Using a different experimental design, Shaun Nichols and Joshua Knobe presented subjects with a description of an alternate universe that is deterministic, and they found that subjects tended to say that agents were not responsible in that universe.

This pattern of conflicting results suggests that subtle features about the way questions about moral responsibility are framed may have an effect upon our intuitions. There are a number of differences between the experiments done by Nahmias and colleagues and those done by Nichols and Knobe. But for present purposes, we are interested in just one of these differences. In some of the experiments by Nahmias and colleagues, the scenario was depicted as holding of our own world, whereas in the experiments by Nichols and Knobe, the scenarios were always set in an alternate universe. For our experiment here, we wanted to see what would happen if we posed questions that differ only in whether the hypothetical situation was set in our own universe or another. On the face of it, one would expect that responses to a hypothetical situation would depend solely upon features of that situation, and not upon the subject’s relation to that situation. Thus, one might expect that responses to questions about freedom and responsibility would be independent of whether those questions were couched in terms of our universe, or another universe just like ours. In the following experiment, we tested this expectation.

---

6 For example, David Hume’s Enquiry concerning Human Understanding (1748).
MORAL RESPONSIBILITY

DOWN TO EARTH

Our study was conducted on University of Utah undergraduates who had not been exposed to philosophical instruction about free will or responsibility.9 This experiment depends upon our ability to convey, in layman’s terms, the essential features of determinism. Although the most precise characterizations of determinism depend on technical philosophical vocabulary about, for instance, laws of nature, we were interested in asking philosophically naïve subjects about their intuitions. Thus, it was important to translate some technical terms into familiar nonphilosophical language. Different characterizations of determinism might produce somewhat different results, but our central interest concerns whether different conditions generate different responses given the same description of determinism. One of the great advantages of contrastive studies is that they can be designed to determine whether one factor makes a difference to a response. The factor of interest to us is whether the scenario is set in the actual world or an alternate universe.

76 participants were randomly assigned to one of two conditions: Actual and Alternate. In both conditions, subjects were given the same sketch of a universe that is regarded as deterministic. In the Actual condition, this universe was clearly implied to be our own:

Many eminent scientists have become convinced that every decision a person makes is completely caused by what happened before the decision—given the past, each decision has to happen the way that it does. These scientists think that a person’s decision is always an inevitable result of their genetic makeup combined with environmental influences. So if a person decides to commit a crime, this can always be explained as a result of past influences. Any individual who had the same genetic makeup and the same environmental influences would have decided exactly the same thing. This is because a person’s decision is always completely caused by what happened in the past.

In the Alternate condition, the universe is explicitly not ours:

Imagine an alternate universe, Universe A, that is much like earth. But in Universe A, many eminent scientists have become convinced that in their universe, every decision a person makes is completely caused by what happened before the decision – given the past, each decision has to happen the way that it does. These scientists think that a person’s de-

9 We hope to explore whether these results extend to other populations. (We hope even more that others will explore whether our results extend to other populations!) However, it is important to note that we are primarily looking at whether subjects from the same population give different answers in the different conditions.
cision is always an inevitable result of their genetic makeup combined with environmental influences. So if a person decides to commit a crime, this can always be explained as a result of past influences. Any individual who had the same genetic makeup and the same environmental influences would have decided exactly the same thing. This is because a person’s decision is always completely caused by what happened in the past.

In both conditions, participants were then asked to rate their level of agreement (from 1 [disagree completely] to 7 [agree completely]) with a statement about the impossibility of moral responsibility in the universe. In the Alternate condition, they were presented with this statement:

If these scientists are right, then it is impossible for a person in Universe A to be fully morally responsible for their actions.

In the Actual condition, participants received the same statement, except that ‘in Universe A’ was removed:

If these scientists are right, then it is impossible for a person to be fully morally responsible for their actions.

The purpose of this experiment was to determine whether judgments would be affected by whether the question was asked of this world or some other possible world. The results were striking: participants in the Alternate condition gave significantly higher levels of agreement than participants in the Actual condition to the claim that it is impossible to be fully morally responsible in that universe.\footnote{The mean response in Actual was 3.58, and the mean response in Alternate was 5.06 (4 is the midline). The difference between the conditions was significant ($t(74) = 3.611, p = .001$).} So, when asked to assume that our own universe is deterministic, people are inclined to judge that people are still morally responsible (in other words, they provide compatibilist responses), but they are inclined to judge of people in another deterministic universe that they are not fully morally responsible, an incompatibilist response.\footnote{Interestingly, these responses parallel Peter van Inwagen’s claim that he thinks that moral responsibility is incompatible with determinism but that if he came to believe determinism was true, he would be a compatibilist (see An Essay on Free Will (New York: Oxford, 1983)). Fisher and Mark Ravizza label this view “metaphysical flipflopping” in Responsibility and Control: A Theory of Moral Responsibility (New York: Cambridge, 1998).}

After answering the question about the impossibility of moral responsibility, participants were asked to indicate their level of agreement with a statement about the moral propriety of blame in the deterministic universe. In the alternate condition, they received the following statement:

Even if these scientists are right, people in Universe A should still be morally blamed for committing crimes.
In the *Actual* condition, participants received the same statement, but 'in Universe A' was removed:

Even if these scientists are right, people should still be morally blamed for committing crimes.

Once again, responses were quite different in the different conditions. Participants in the *actual* condition gave significantly higher levels of agreement to this claim. That is, they were significantly more likely than those in the *alternate* condition to give compatibilist responses to this question as well.\(^{12}\)

Finally, participants were also asked to indicate their level of agreement with a statement about the existence of free choice in the deterministic universe. In the *Alternate* universe, they received the following statement:

If these scientists are right, then it is impossible for people in Universe A to make truly free choices.

In the *Actual* condition, participants received this:

If these scientists are right, then it is impossible for people to make truly free choices.

For this question as well, there was a significant difference between conditions. Participants in the *Alternate* condition gave higher ratings of agreement to this statement than those in the *Actual* condition.\(^{13}\)

**IMPLICATIONS FOR OUR UNDERSTANDING OF MORAL RESPONSIBILITY**

The above results have several implications for our understanding of moral responsibility. First, they provide one way of interpreting some of the apparently conflicting findings of Nichols and Knobe and of Nahmias and colleagues (*op. cit.*). For we find that subtle differences in the way questions about moral responsibility are framed can make

\(^{12}\) The mean response in *Actual* was 5.35, and the mean response in *Alternate* was 3.67. The difference between the conditions was significant ($t (65) = -4.426, p < .001$). Note that the question was positively framed, while the other two were framed negatively, so we in fact would expect this sort of inversion in the agreement values. These values are consistent with what would be expected if judgments of moral responsibility and blame-worthiness were correlated.

\(^{13}\) While the difference is significant ($t (74) = 2.362, p < .05$), in both conditions, the mean response was above the midline. In the *Actual* condition, the mean response was 4.3, and in the *Alternate* condition, the mean response was 5.3. Overall, these results suggest that there was stronger agreement to the statement that people (in a deterministic universe) lacked freedom of choice than that they lacked moral responsibility. Interestingly, we found that this difference was statistically significant in the *Actual* condition ($t (39) = -2.456, p < .05$), but not in the *Alternate* condition ($t (35) = -.843, p = .405$, n.s.).
This graph shows average ratings for level of agreement with statements about the moral responsibility, blameworthiness, and freedom of agents in deterministic worlds if those worlds are the actual world or a world in some alternate universe. Ratings of 1 correspond to disagree completely, 7 with agree completely, and 4 is neutral. Level of agreement was assessed to the following questions: in such a world (1) it is impossible for a person to be fully morally responsible for their actions; (2) people should still be morally blamed for committing crimes; and (3) it is impossible for people to make truly free choices.

a difference to the intuitions that are elicited. In some of the scenarios from Nahmias and colleagues, subjects are explicitly asked to assume that determinism is true of our universe, whereas in Nichols and Knobe’s scenarios, subjects are always asked to assume that determinism is true of an alternate universe. Thus, our results suggest that part of the explanation for the diverging responses might be the setting of the scenario. When the deterministic universe is our own, people are more likely to give compatibilist responses.

Second, the results bear upon an issue that has captured both the philosophical and the popular imagination. Some philosophers and many laypeople fear that catastrophe will follow if people come to accept determinism. For example, Saul Smilansky suggests that if people come to realize the absence of indeterminist choice, “our fundamental values, practices, and attitudes, such as abhorrence about the ‘punishment’ of the innocent, the inherent value we put on ‘equality of opportunity’, belief in our potential for blameworthiness ... are very likely to be harmed.” Smilansky worries that people “might succumb

to ... an unprincipled nihilism” (ibid., p. 189). As a result, although Smilansky thinks that free will is an illusion, he maintains that we should not try to dispel the illusion. Related claims have been made in the popular press in the context of neuroethics. Among the neuroethical worries raised by technological advances in neuroscience is that our improving scientific understanding of higher brain functions will cause the public to view currently unexplained psychological phenomena such as choice and decision-making as a merely mechanical processes, and that people will come to believe that human action is merely the result of mechanism in a deterministic universe. The projected upshot of this potential change in belief is that upon coming to see us as merely mechanisms in a deterministic world we will come to realize that we lack free will, and consequently, moral responsibility. This in turn, the argument continues, will undermine the moral fabric of our society; the chaos that will result is left to our imaginations.

This line of reasoning is often meant to cast doubt on the ethics of pursuing neuroscientific research into the brain bases of higher cognition. This ploy is questionable for a number of reasons. What is crucial for our purposes, however, is whether the folk will stop treating people as morally responsible if determinism becomes widely accepted. Our experiment addresses this concern in two ways. First, it provides some explanation for the expectation that the belief in determinism would lead us to abandon our belief in moral responsibility. For when asked about a hypothetical situation in which determinism is true (the Alternate case), people are inclined to claim that moral responsibility is excluded. More importantly, though, the experiment suggests that the practical worries are misplaced, for our judgments about moral responsibility, should we come to believe determinism to be true of the actual world, would probably not be unseated by this

15 For a discussion of why such a conclusion is not warranted on the basis of neuroscientific research, see Adina L. Roskies, “Neuroscientific Challenges to Free Will and Responsibility,” Trends in Cognitive Sciences, x (2006): 419–23.

beliefs.\(^\text{17}\) The upshot of this is that these worries about how neuroscientific understanding will undermine the social order are misplaced.\(^\text{18}\)

Other philosophers agree with Smilansky that people’s attitudes towards responsibility will change markedly if they come to believe in determinism, but instead they hail this as a much needed revolution. Thus, Joshua Greene and Jonathan Cohen write, “As more and more scientific facts come in, providing increasingly vivid illustrations of what the human mind is really like, more and more people will develop moral intuitions that are at odds with our current moral practices.”\(^\text{19}\)

In particular, they maintain, we will stop thinking that people are responsible and that the guilty deserve punishment. “The law will continue to punish misdeeds, as it must for practical reasons, but the idea of distinguishing the truly, deeply guilty from those who are merely victims of neuronal circumstances will ... seem pointless” (ibid.). Our results suggest that we should expect neither revolution, as do Greene and Cohen, nor catastrophe, as do Smilansky and the naysayers in the popular press. If people came to believe in determinism, it seems likely that they would not significantly change their practices of attributing responsibility.

**WHAT IS GOING ON?**

Our results are surprising—Why should intuitions about moral responsibility depend upon factors other than the basic facts about the world? There are several models by which to understand these findings.

**Depth of Processing.** One explanation of our results appeals to differences in the depth of processing in the different conditions. It is an old and venerable view in social psychology that people process problems more fully and accurately when the questions are personally relevant.\(^\text{20}\) For instance, in one study, undergraduate participants were


\(^{18}\) This is not to say that neuroscientific advances pose no threat to the social order; only that this particular focus for worry is overblown.
