**Folk Fears about Freedom and Responsibility:**
**Determinism vs. Reductionism**

*The Journal of Cognition and Culture* vol. 6 (1-2), pp. 215-237
** Please cite and quote from published version **

Eddy Nahmias
Department of Philosophy
Brains & Behavior Program
Georgia State University
Atlanta, GA 30302-4089
enahmias@gsu.edu

## 1. Folk Intuitions and Folk Psychology

My initial work, with collaborators Stephen Morris, Thomas Nadelhoffer, and Jason Turner (2005, 2006), on surveying folk intuitions about free will and moral responsibility was designed primarily to test a common claim in the philosophical debates: that ordinary people see an obvious conflict between determinism and both free will and moral responsibility, and hence, the burden is on compatibilists to motivate their theory in a way that explains away or overcomes this intuitive support for incompatibilism. The evidence, if any, offered by philosophers to support the claim that incompatibilism is intuitive has consisted of reports of their *own* intuitions or informal polls of their students. We were skeptical about the reliability of such evidence, so we used the methodology—now associated with the label 'experimental philosophy'—of conducting formal surveys on non-philosophers. Our participants read a scenario that describes a deterministic universe and were then asked to judge whether agents in those scenarios act of their own free will and are morally responsible for their actions. Using three different scenarios with hundreds of participants, we consistently found that the majority (2/3 to 4/5) responded that agents in deterministic universes *do* act of their own free will and *are* morally responsible. That is, we found that most ordinary folk do *not* seem to find incompatibilism intuitive or obviously correct.

Our results have been challenged in various ways, philosophical and methodological. For instance, Shaun Nichols (2004, this volume) and Nichols and Joshua Knobe (unpublished) offer some experimental evidence suggesting that, in certain conditions, most people express incompatibilist and libertarian intuitions. I will respond to this work in the following section. I agree that people express conflicting intuitions about free will (after all, we consistently found a minority of participants expressing incompatibilist intuitions), but in section 3 I will offer an alternative account of what drives these conflicts.

However, Nichols and I agree more than we disagree. First, we agree that the methodology of experimentally surveying non-philosophers' judgments plays an important role in what he aptly calls the *descriptive* project of mapping ordinary intuitions and conceptual usage. And this descriptive project is a necessary first step in both the *substantive* and *prescriptive* projects of determining in what ways, if any, these folk intuitions and concepts are mistaken and whether they should be accordingly revised.[1] I also appreciate that Nichols has, as

[1] Notice that, contrary to some critics of experimental philosophy, neither Nichols nor I have suggested that discovering what the folk intuitions are (where that is possible) plays any decisive role in the substantive philosophical project. At a minimum, however, this methodology raises interesting meta-philosophical questions about the nature of intuitions and their role in conceptual analysis and theory building, conceptual revision, burden

it were, taken this methodology to the next level—to use it as an initial step in understanding the psychological mechanisms that generate people's intuitions and judgments regarding philosophical issues, such as free will and moral responsibility. This research project has the potential to help us understand (a) why people believe what they do regarding issues relevant to philosophical debates, (b) why people sometimes express conflicting beliefs, (c) why some mechanisms that generate beliefs more reliably track the truth (or at least generate more consistent beliefs) than other mechanisms, and potentially (d) why certain philosophical debates may derive in part from conflicting intuitions generated by competing psychological mechanisms. Evidence about folk *intuitions* can thereby serve as evidence about folk *psychology*.

In this paper I will offer some alternative interpretations of Nichols' experimental results and then present new experiments whose results suggest alternative psychological mechanisms are driving some of the conflicting results of folk surveys. My goal is to bolster the claim that most people's pre-philosophical intuitions do not pick out *determinism* as a threat to free will and moral responsibility, but instead ordinary people fear that certain reductionistic or mechanistic descriptions of decision-making and action conflict with freedom and responsibility. I believe that intuitive support for incompatibilism derives in large part from presenting determinism in a way that co-opts these fears of reductionism. Put simply: the most intuitive pre-philosophical view is that free will and responsibility are possible so long as our conscious deliberations, plans, and decisions make the right sort of causal contribution to our actions, so any thesis that suggests our actions are caused by processes that bypass our conscious mental life—that bypass *us*, as it were—is intuitively threatening to freedom and responsibility; determinism *per se* is generally not recognized to be such a 'bypass' threat—nor do I think it should be—whereas certain reductionistic views are naturally seen as such 'bypass' threats. I will conclude (section 4) with a brief discussion of the various conflicting intuitions the nascent experimental work has uncovered, conflicts that potentially help to explain the enduring nature of philosophical debates about free will.

**2. Alternatives to Incompatibilist Interpretations**

Nichols rightly points out that it is a messy and difficult business to map folk intuitions about complex philosophical issues (I won't rehash here the arguments for engaging in such a difficult task in the first place). There are numerous methodological difficulties, including two particularly thorny ones: probing the folks' modal intuitions and ensuring they reason counterfactually. So, I raise the following questions about Nichols' experiments and interpretations with no lack of respect for his innovative attempts to elicit the relevant intuitions and with no confidence that my alternative interpretations are the final word on the matter.

Nichols describes some experiments that suggest ordinary folk (both children and adults) tend to think physical events are deterministic whereas human choices are indeterministic.[2] First, note that most compatibilists have no stake in whether human choices are deterministic or not; if choices are in fact indeterministically caused, that does not entail that they would be unfree if they *were* deterministically caused. For many compatibilists (like me), the point is that whether

---

of proof, error theories, etc. Furthermore, those who think the methodology is *entirely* irrelevant to philosophical debates should either offer a better way to determine what is commonsensical or intuitive or urge philosophers to stop making claims about what is commonsensical or intuitive to ordinary folk.

[2] See his section 3.1 (this volume) and his 2004, where he argues that children conceive of agency in a way suggestive of agent-causal theories. See Turner and Nahmias (forthcoming) for more detailed discussion of some of the issues discussed here.

determinism or indeterminism is true is *irrelevant* to free will and moral responsibility—what matters is the particular way our choices come about, not whether there was some probability they would occur differently.[3] So, if the folk theory is that choices are indeterministic, that would only support the idea that incompatibilism is intuitive to the folk *if* they have a libertarian conception of agency—that is, if they believe that our choices being indeterministic is *required* for (and what allows for) freedom and responsibility. Nichols interprets his evidence in this way. I suggest an alternative interpretation for at least a significant number of people's responses (i.e., I don't deny that *some* people seem to have pre-philosophical libertarian intuitions).

When asked whether a physical event (e.g., water coming to boil) *has* to occur given the same prior conditions, Nichols found that most participants said *yes*, but when asked whether a moral choice (e.g., to steal a candy bar) has to occur given the same prior conditions (and desires), most said *no*. Nichols' takes this difference in responses to suggest that people are libertarians about human choice, because they think human choices are uniquely indeterministic. Instead, I think it is likely many people simply respond as indeterminists about certain *complex* processes. For simple processes, such as water boiling, holding fixed prior events may be considered sufficient to ensure the culminating event, but for complex processes, such as the weather, holding fixed prior events may *not* be considered sufficient to ensure later events. Some human decisions may be seen as complex in this sense and this might explain the pattern of responses Nichols got.

I ran an experiment to test this hypothesis. Ninety-nine participants (college students) read the three scenarios below (counterbalanced for order) and then answered one question about each, circling either 'Yes', 'No', or 'I don't know'. They had an opportunity to explain their answers on the back of the survey.

> **Scenario L**: Imagine that a lightning bolt hits a particular tree at a particular time. Now imagine that the universe is re-created over and over again, starting from the exact same initial conditions (and with all the same laws of nature). If that were the case, do you think that every time the universe is re-created everything would happen the exact same way, including the lightning bolt's hitting the tree at that time?
>
> **Scenario I**: Imagine a woman is trying to decide between ordering vanilla ice cream and chocolate ice cream, and at a particular time she decides on vanilla. Now imagine that the universe is re-created over and over again, starting from the exact same initial conditions (and with all the same laws of nature). If that were the case, do you think that every time the universe is re-created everything would happen the exact same way, including the woman's deciding to order vanilla at that time?
>
> **Scenario S**: Imagine a woman is trying to decide whether or not to steal a necklace, and at a particular time she decides to steal it. Now imagine that the universe is re-created over and over again, starting from the exact same initial conditions (and with all the same laws of nature). If that were the case, do you think that every time the universe is re-created everything would happen the exact same way, including the woman's deciding to steal the necklace at that time?

[3] Rampant indeterminism could be problematic if it undermined the requisite regularity between, for instance, one's reasons and choices or one's intentions and actions. 'Soft determinists' were compatibilists who argued that determinism is necessary for free will, but few compatibilists seem to hold that view anymore.

Results suggest conflicting beliefs among participants about whether the events would happen every time, though consistently across all *three* scenarios more participants responded that the events would *not* happen the same every time the universe was re-created than responded that the events would happen the same way every time.[4]

| | Yes | No | I don't know |
|---|---|---|---|
| Lightning (L) | 42 | 49 | 8 |
| Ice Cream (I) | 36 | 52 | 11 |
| Steal necklace (S) | 36 | 55 | 8 |

Of the 99 participants, 30 were 'determinists' who answered 'yes' to all three scenarios, 40 were 'indeterminists' who answered 'no' to all three, and 29 were 'complex cases' who offered different responses to the scenarios. Most of the determinists offered explanations suggesting that they believed that the same conditions and laws must produce the same outcome, and that this is true of human choices too (e.g., 'Certain things happen as a result of what happened before it. If the situations were recreated exactly, then there would be no other choice but for [the] occurrence to happen').[5] Of the indeterminists, about half offered explanations that suggested there is some randomness or chaos in the universe, without suggesting any difference between decision-making and natural processes (e.g., 'In a universe where things happen randomly all the time why, if reversed, would those same random things happen the exact same way?'). Of the remaining indeterminists and the 'complex cases,' about 20 participants made reference to something specific regarding humans, with 13 suggesting that humans have choice or free will so that they can produce different outcomes (e.g., 'Humans may be able to change their thinking and their ways. The lightning strike doesn't fit into this because it was inevitable'). Overall, only 9 of 99 participants responded that the lightning bolt would happen the same way every time but that the human choices would *not*. Hence, these results suggest that, at least on this sort of probe, (a) some people view choices in a deterministic way (which is also supported by the data Nichols presents in section 3.2), (b) few people draw a distinction between physical processes and human choices with respect to determinism vs. indeterminism, and (c) the complexity of a causal process may drive indeterminist responses at least as much as any libertarian intuitions people may have.[6]

Nichols offers other evidence that people have an incompatibilist conception of choice. He points out that, in his work with Knobe, the vast majority of people responded that an indeterministic universe, B, where every event, *except* for human choices, is completely caused by prior events, is more like our universe than a deterministic one, A, where every event is

[4] Since there happened to be 99 participants, the absolute numbers in the table can also be treated as percentages. No order effects were found.

[5] These 'determinists' did not, however, suggest that they thought the choices were thereby unfree. I always offer participants an opportunity to explain their answers on surveys and, like Nichols, I have found that these explanations often offer helpful insight into folk intuitions and theories.

[6] The preceding two paragraphs are drawn largely from Turner and Nahmias (forthcoming). One alternative interpretation of these results is that even in the lightning scenario some people are expressing libertarian intuitions about choice rather than indeterminist intuitions about complex physical processes: it asks whether *everything* would happen the same way, so if someone reads that to include (unmentioned) human choices, they may respond 'No' because they think (at least) some human choices would *not* happen the same way every time, even if they think the lightning bolt would.

completely caused by prior events.[7] I'm concerned, however, that the differences in the 'completeness' of the causation in the universes is not what most participants are responding to. I suspect that what is more salient to them is the different way human decisions are described in the two universes. Participants read descriptions that explicitly say of Universe B, "She could have decided to have something different" (while the description of A leaves out this sentence) and that conclude, "in Universe B … each human decision *does **not** have to happen* the way it does." When looking for features that make one universe more like ours than the other, as the first experimental question demands, it would be easy to read these phrases to suggest the opposite is true of Universe A—that humans could *not* decide to do anything different and that each human decision *has to happen* the way it does. If participants are interpreting the scenarios in this way, then their answers cannot inform us about whether they think free choices are or can be deterministically caused.

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 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). This is because the prior events that caused *X* did not *have to* happen (nor, for that matter, did determinism *have to* be true). Most philosophers agree that even in a deterministic universe, natural events are still contingent; because past events (and laws of nature) are not necessary, neither are the events they deterministically cause. What philosophers disagree about is whether determinism entails that agents do not have the requisite ability to do otherwise to be free and responsible, as suggested for instance by Peter van Inwagen's Consequence argument (1983). Since this question about the ability to do otherwise is one of the questions at the heart of the debate between compatibilists and incompatibilists, then *if* either side cares whether its premises and principles are pre-philosophically intuitive, surveys of the folk should not describe determinism in a way that begs the question by suggesting it entails that every event (including every decision) that happens *must* happen the way it actually happens.[8]

Now, I recognize that Nichols and Knobe's scenarios do not make exactly this modal error: when describing Universe A, they write, "<u>given the past</u>, each decision *has to happen* the way that it does" (I've underlined the crucial phrase). My point here is that the language I quoted earlier invites the modal error—participants are asked to contrast Universe A with one where "each human decision *does **not** have to happen* the way it does." If they *interpret* this to mean that Universe A is one where each human decision ***does** have to happen* the way it does, then it would be expected that they would think that this universe was both unlike ours and one where people could not be free and responsible. But it would also not inform us about whether they thought determinism, properly understood, precludes free will.

Notice also that the description of Universe A does make a different, more subtle modal error. It says "given the past, each decision *has to happen* the way that it does" and "if everything in this universe was exactly the same up until John made his decision, then it *had to*

---

[7] There are complex questions about the relations between complete causation, determinism, and indeterminism. For instance, indeterministic quantum events may be completely caused. Here is a place where surveying folk intuitions requires simplifying the philosophical issues, perhaps in a problematic way.

[8] Determinism entails that □[(Po & L) ⊃ P]—i.e., necessarily, *given* the actual past state of affairs (Po) and the actual laws of nature (L), there is only one possible present state of affairs (P). But determinism does *not* entail (fatalism) that □P (or that □Po or □L)—i.e., that the actual state of affairs (or the actual past or laws) are necessary (could not be otherwise).

*happen* that John would decide to have French Fries." These statements are most accurately read as saying, "Holding fixed the actual past causal conditions [and presumably the laws too], then *necessarily* subsequent events occur as they actually do."[9] But according to any conditional analysis of the ability to do otherwise, subsequent events *cannot* happen otherwise *if* past conditions are held fixed—that is, later events can be different *only if* past conditions had been different. Such analyses require relevant past conditions to be different in order for present conditions to be different. A compatibilist, for instance, might believe that we can do otherwise in the sense relevant to freedom and responsibility as long as we would do otherwise given relevantly different past conditions or laws. Such a compatibilist would think we *cannot* do otherwise in the relevant sense if past conditions are held fixed.

Obviously this complaint looks highly esoteric and perhaps too removed from anything relevant to understanding folk intuitions. But the worry is that describing determinism in a way that suggests we could not have decided otherwise will skew the results towards libertarian and incompatibilist responses; and I worry that the description of Universe A (contrasted with Universe B) suggests that *nothing*, including decisions, could happen otherwise than they actually do. Consider Nichols' interesting evidence (presented in section 3.2) suggesting that people often seem to use a deterministic theory to explain and predict human decisions. If so, it would not be surprising if people sometimes understood decision-making (and other events) in a conditional way: e.g., If she decided to *X*, then it's because prior conditions *a, b, c…* occurred, but if she had decided to *Y*, then it would have been because prior conditions *p, q, r…* would have occurred. A scenario that suggests that the decision to *X* ***had*** *to happen*—either *simpliciter* or because *a, b, c…had* to happen—would conflict with this understanding of determinism and the ability to do otherwise.

Now, I'm certainly not confident that any further experimental work would bear out my alternative interpretations to Nichols' incompatibilist ones. In part, this is because the modal issues I've highlighted are so complex. One of the central issues in the free will debate is, *descriptively*, how people in fact interpret the ability to do otherwise required for freedom and responsibility and, *substantively*, how they *should* interpret it. These issues are inextricably tied up with intuitions about necessity and possibility, and such intuitions twist *philosophers* into knots.[10] But, given the concerns I have raised, I do not (yet) concede that Nichols has found evidence that most ordinary people, in relevant contexts, express intuitions properly interpreted as libertarian or incompatibilist.

In any case, Nichols and Knobe's most interesting finding was about one particular contextual factor. They found a significant difference between participants' responses to questions about moral responsibility in an *affect-neutral* condition vs. their responses in an *affect-laden* condition. And that's Nichols main point—that people have conflicting intuitions about agency and responsibility that get triggered by specific factors. This is a fantastic

---

[9] That is, they suggest that determinism entails that [(Po & L) ⊃ □P] rather than the proper □[(Po & L) ⊃ P]. The latter would be written, "Necessarily, holding fixed the actual past (and laws), then subsequent events occur as they actually do," or in the authors' terms: "It *has to happen* that, given the past, each decision happens the way it does."

[10] In our original studies, my collaborators and I (2005) asked our participants about the ability of agent's to do (or choose) otherwise in the three deterministic scenarios we developed and the results were pretty messy—often around 50-50 with few consistent connections between answers to this question and answers to questions about the agents' free will or moral responsibility. An alternative methodology to probing naïve participants' modal intuitions is to tutor students on the modal concepts (and other relevant issues such as laws of nature, states of affairs, entailment, etc.) without introducing any issues regarding freedom and determinism, and then probe their intuitions about these issues.

discovery because it forces philosophers to explain which factors properly evoke, and which factors illegitimately distort, any intuitions relevant to developing a philosophical theory or analyzing a concept. I suspect that the authors are correct that situations that prime certain emotions influence judgments about whether determinism precludes responsibility. (If I had to lay my cards on the table, I'd take the Strawsonian line that responsibility attributions properly derive from our reactive attitudes and should be generalized from judgments about specific agents acting in specific cases, many of which will evoke emotional responses in us, which is *not* to say that such attributions are never improperly distorted by emotions).

I have no significant quibbles about Nichols and Knobe's discovery that affect plays a significant role in intuitions about responsibility.[11] Instead, I will now turn to a different factor that seems to have a significant effect on people's judgments about freedom and responsibility.

**3. An Experiment on Folk Fears of Determinism vs. Reductionism**

Consider the way some philosophers have presented the consequences of determinism for freedom and responsibility:

> An agent would not be morally responsible at all if he was caused necessarily, predetermined, to try to do what he did, by his brain state, and that in turn by some prior state, until we come to causes outside the agent's body and ultimately to causes long before his birth. (Richard Swinburne quoted in Fischer 1994, p. 6)
>
> What am I but a helpless product of nature, destined by her to do whatever I do and to become whatever I become? (Richard Taylor 1963, p. 36)
>
> [Determinism means that our] self-monitoring and self-critical capacities, so essential to human nature, might as well dry up and wither; they would no longer have any function. (Joel Feinberg quoted in Fischer 1994, p. 4)

[11] I do have some minor quibbles. I worry that some of the difference in results that they attribute to affect might be explained by participants considering different comparison classes in the affect-neutral vs. affect-salient cases. In the affect-neutral case, participants are asked to consider whether agents could be "fully morally responsible" in universe A and then asked about agents in universe B. If participants feel any demand conditions to find some difference between the scenarios, then given the issues I have raised about the description of universe A, it is not surprising that universe B is considered both more like ours and more amenable to *full* moral responsibility than universe A (some participants may be thinking that one of the universes must allow more responsibility than the other one, though both may allow some). In the affect-salient cases, however, participants are asked only about whether an agent (Bill), who does a *very* bad thing (killing his wife and children!) in universe A, could be fully morally responsible. In this case, participants may be more likely to compare not universe A with universe B but rather *Bill* with other agents in his world. In that case, since no information is given to distinguish Bill from other agents in his world, some participants' might be expressing their belief that there is no reason Bill should be considered any *less* responsible than others, no doubt influenced by their emotional response to Bill's horrendous deed. Furthermore, some participants may interpret the question about moral responsibility to mean, "Should Bill be punished for his action?" and even if they do *not* think he has free will, they may think he needs to be punished. This interpretation would not be primed in the affect-neutral case since there is no specific action to be (potentially) punished.

Though I have agreed that affect likely plays a role in intuitions about responsibility, I should point out that in our own studies we found no significant differences in participants' judgments about the freedom or moral responsibility of agents acting in deterministic universes when those agents were doing morally bad actions (e.g., robbing a bank, stealing a wallet), morally good actions (e.g., saving a child, returning a wallet), or morally neutral actions (e.g., going jogging). See Nahmias *et al*. (2005, pp. 567-571).

> In those rides that amusement parks sometimes provide, in which one sits in a car that follows a track through some darkened room of illuminated objects, the car sometimes has a steering wheel. If one turns the wheel in the directions suggested by the environment—directions in which the car is actually going—one can easily get the feeling that one is steering the car—even though one knows all along that he is not. A child might think he actually was steering the car. (Carl Ginet quoted in Fischer 1994, p. 14)

Of course, these authors (and others) also present arguments for incompatibilism to supplement these rhetorical presentations of determinism. But when it comes to pumping intuitions so that people being introduced to the philosophical debate feel the pull of incompatibilism, I suspect these sorts of presentations are doing a lot of the work. And they do so illegitimately. They suggest that determinism entails that our brain states completely cause us to do what we do, that the forces of nature coerce us, that our self-reflective capacities are causally irrelevant, and that we are on a predetermined track such that our experience of steering the course of our lives is illusory. But determinism does not entail fatalism (the view that some things will happen no matter what we do), nor coercion by natural forces (or by the past or the laws of nature), nor reductionism or epiphenomenalism (e.g., the view that our brain states cause everything we do while our conscious mental processes—including our self-critical capacities—play no causal role in our decisions or actions). Fatalism, coercion, reductionism and epiphenomenalism are each threatening—at least intuitively—to free will and responsibility, but it would take an impressive argument to show that determinism entails any one of them. The reason each of these views are threatening is not because the past and laws are sufficient for our behavior but because they suggest that our behavior is caused by forces that *bypass* our conscious mental life—or at best, by forces that manipulate our conscious mental life.[12]

For my purposes here, I am not trying to show that incompatibilist arguments work by conflating determinism with these other threats. Rather, I only hope to show that ordinary people respond differently to deterministic situations that evoke a reductionistic picture of decision-making that suggests such bypassing as compared to deterministic situations that do not suggest bypassing because the agents' psychological states play a role in their decisions and actions. Such a difference in people's responses to two different ways of presenting determinism would support the idea that ordinary folks' *apparent* incompatibilist intuitions may derive from

---

[12] I don't mean to suggest that *all* forms of reductionism should be considered a threat to free will. I am referring to the sorts of reductionism that might suggest epiphenomenalism.

I also don't mean to suggest that *compatibilists* are less likely to pump certain intuitions to bolster their arguments—consider Frankfurt cases—and that some such pumping is not illicit. If most people have conflicting intuitions about free will and responsibility, then philosophers may be particularly adept at developing thought experiments or arguments that accentuate one set of conflicting intuitions. The deflationary worry would be that certain philosophical debates linger because philosophers' own intuitions become accentuated in one direction and then solidified in support of one theory.

In Kane (2003), he suggests that compatibilists subvert "ordinary persons' natural incompatibilist" intuitions by pumping the intuition that indeterminism cannot help secure free will because it entails randomness, luck, lack of control, etc. My argument might be seen as taking a similar form: incompatibilists subvert ordinary person's natural *compatibilist* intuitions by pumping the intuition that determinism entails reductionism, epiphenomenalism, coercion by natural laws, etc.

intuitions that reductionism is threatening to freedom and responsibility rather than determinism *per se*.[13]

I will mention, however, that I think even sophisticated incompatibilist arguments sometimes smuggle in these reductionistic pictures to support the intuitive plausibility of their premises and principles. In addition to the examples cited above, consider Peter van Inwagen's Consequence argument. He defines determinism in terms of "the state of the entire *physical* world" and "the laws of *physics*" (1975, p. 47, my italics), and elsewhere he adds that "the laws of nature would be just as they are even if there had never been any human beings or other rational animals" and he stipulates that "*psychological* laws" not be included in the conception of laws of nature to be used in the argument (1983, pp. 60-64). When van Inwagen says that it is obvious that one of his crucial premises (NL) is true—that no one has a choice about what the laws of nature are—it helps his case that, by stipulation, the laws not include any psychological laws that might refer to actual human deliberations, decisions, and actions.[14]

It may be true that psychological laws are reducible to the laws of physics such that psychological states or properties play no genuine causal role in producing our actions, but it is not clear how this conclusion follows simply from the truth of determinism rather than from an argument that does not require determinism at all—e.g., the causal exclusion argument (Kim 1998). If determinism is consistent with the laws of psychology being irreducible and with psychological states and properties having genuine causal influence on our actions—and I see no reason why it is not—then incompatibilist arguments should not rely on the intuitive appeal of premises that suggest otherwise. If determinism *does* entail a type of reductionism that allows no causal role for mental states—though I see no reason why it should—then I think that the argument for incompatibilism would be much stronger. But I think its strength would actually derive from the threat of epiphenomenalism rather than the logical sufficiency of the past and laws for our actions. One way to see this is to notice that if reductionism and epiphenomenalism are equally entailed by *indeterminism* (say, of the quantum mechanics variety), then free will and responsibility may appear equally threatened, suggesting that it is not determinism that is doing the work.

But enough of the *substantive* project. Back to the *descriptive* project. My hypothesis is that a principal psychological mechanism that drives incompatibilist intuitions involves people's fear of reductionistic descriptions of deliberation and decision-making. While some people do seem to express a 'pure' incompatibilist intuition that any choice that has prior sufficient causes cannot be free or deserving of praise or blame, I think most people do not have this intuition and are not threatened by the possibility (or actuality) of 'psychological determinism.' In addition to my group's original studies where the deterministic scenarios did not rule out mental causation and where most participants expressed compatibilist-friendly intuitions, notice also that Nichols reports two studies (in his section 3.2) in which most people's responses indicate a belief in psychological determinism. Since everyone's studies have shown that nearly all the folk believe humans are free and responsible, this belief in psychological determinism might suggest that people do not think it precludes freedom and responsibility. It's true that one way to reconcile

---

[13] Another way to demonstrate this claim is to present a reductionistic picture as deterministic to some people and as *indeterministic* to others. I predict there would be no significant difference in people's responses about agents' free will and responsibility in such scenarios, suggesting that it is not determinism but reductionism that is at issue.

[14] David Lewis' (1981) response to van Inwagen (1975)—that, even if determinism is true, we can sometimes do otherwise such that *if* we did otherwise, some actual law would not have been a law—does not rely on there being non-reductionistic psychological laws, but I think this possibility makes Lewis' response more intuitively plausible.

these findings is to say, instead, that people have conflicting intuitions or folk theories—a libertarian conception of freedom and responsibility combined with a mindreading system that conceives of agency as deterministic—and then to argue that most people simply don't recognize the conflict between these conceptions, at least not until philosophers point it out to them.

But another way to interpret the data is that most people do *not* see a conflict between their strong belief that we are free and responsible and their deterministic folk psychology, because their folk psychology is inherently non-reductionistic, explicitly requiring a role for conscious beliefs, desires, reasons, plans, and deliberations to cause our choices and actions.[15] While this type of theory would see no conflict between responsibility and certain types of determinism, it would recognize a threat to free will and responsibility from certain *types* of causal explanation—specifically, ones that suggest the mentalistic components of decision-making, such as our conscious beliefs, desires, goals, plans, reasons, etc.—are not causally relevant to our decisions and actions. Since it is very hard (even for philosophers!) to reconcile the causal relevance of mental states with a reductionistic and mechanistic picture—for instance, one that describes decision-making and action in terms of neurobiological processes—it would not be surprising if most people found such a 'neuro-reductionistic' picture threatening to free will and responsibility. Indeed, I suspect that the claim that incompatibilism is intuitive to ordinary people rests on a failure to distinguish 'pure' incompatibilism (between determinism *per se* and free will) and 'derivative' incompatibilism (between deterministic *reductionism* and free will).

To test this hypothesis I developed two scenarios, both of which describe agents' decisions and actions as being completely caused by prior events, but that differ with respect to the type of events that cause the decisions and actions. Because earlier work has indicated some problems with asking people to reason counterfactually about scenarios describing humans, I used a 'twin earth' type set-up. Each participant read one of the following two scenarios (the only differences in wording are underlined)[16]:

> **Scenario**: Imagine there is another universe similar to ours, in which there is a planet, named Erta, similar to ours in many ways. The landscape and life there look much like Earth, and there are advanced life forms (Ertans) who look, talk, and behave much like we do. However, the Ertans' science has advanced far beyond ours. Specifically, the Ertan neuroscientists have discovered exactly how Ertans' brains work. The neuroscientists have discovered that every single decision and action Ertans perform is *completely caused* by the particular chemical reactions and neurological processes occurring in their brain at the time, and that these chemical reactions and neurological processes in the brain are *completely caused* by earlier events involving their particular genetic makeup and physical environment. So, whenever Ertans act, their action is completely caused by the particular chemical reactions and neurological processes occurring in their brain at the time, and these brain processes are completely caused by

[15] Notice that Nichols' studies in section 3.2 use scenarios that describe the psychology of the agents as both deterministic and *non*-reductionistic, at least in the sense that the psychological states are not identical to the underlying physical states but multiply realizable.

[16] Participants were Georgia State University students enrolled in Critical Thinking classes. Results reported below are based on those 49 participants in my sample who had not taken a college philosophy course and who correctly answered the manipulation checks.

> earlier events that trace back to their particular genetic makeup and physical environment.
>
> **Scenario**: Imagine there is another universe similar to ours, in which there is a planet, named Erta, similar to ours in many ways. The landscape and life there look much like Earth, and there are advanced life forms (Ertans) who look, talk, and behave much like we do. However, the Ertans' science has advanced far beyond ours. Specifically, the Ertan psychologists have discovered exactly how Ertans' minds work. The psychologists have discovered that every single decision and action Ertans perform is *completely caused* by the particular thoughts, desires, and plans they have at the time, and that these thoughts, desires, and plans are *completely caused* by earlier events involving their particular genetic makeup and upbringing. So, whenever Ertans act, their action is completely caused by the particular thoughts, desires, and plans they have at the time, and these thoughts, desires, and plans are completely caused by earlier events that trace back to their particular genetic makeup and upbringing.

Participants then circled either "Yes," "No" or "I don't know" to two experimental questions: (1) Now pretend that the scenario above is true and it accurately describes the Ertans. Assuming that is the case: Do you think that when the Ertans act, they can act of their own free will? (2) Do you think that Ertans deserve to be given credit or blame for their actions?

Of the participants who read the neuro-reductionistic determinism (ND) scenario, only 18% (4 of 22) responded that the Ertans act of their own free will and only 19% (4 of 21) responded that they deserve credit or blame for their actions. But of the participants who read the psychological (nonreductionistic) determinism (PD) scenario, 72% (18 of 25) responded that the Ertans act of their own free will and 77% (17 of 22) responded that they deserve credit or blame for their actions (see Figure 1).[17]

[INSERT FIGURE 1]

[17] Most participants answered the two questions consistently—discrepancies in totals are from the few who answered "I don't know" to one of the questions. After answering these questions, participants then turned the survey over to answer the manipulation checks, to explain their answers, to say whether they think scientists on Earth will discover the same is true of us, and to offer some demographic information. The manipulation checks aim to ensure that participants answered the experimental questions with an understanding of the complete causation described in the scenarios.

A methodological tangent: Nichols and I agree that, when surveying the folk, it is important to consider both the explanations participants offer for their answers and the responses of those who miss the manipulation checks. It seems that some participants who miss manipulation checks are not reasoning counterfactually—they are not pretending the scenario is true when they offer their answers, but instead they seem to be answering the questions based on what they think is true (e.g., that we are free and responsible). Then they miss the manipulation checks precisely because they think the scenario cannot be accurate as described. Such responses can be revealing. Consider, for instance, that in my experiment there were no significant differences in responses between those who passed and those who failed the manipulation checks—except in one case: in the ND scenario, roughly a third of participants missed the manipulation check, and of these, 75% and 77% responded that the Ertans were, respectively, free and responsible (vs. 18% and 19% who responded that way and then passed the checks.) Their explanations often refer to their belief that the Ertans are like us, and we are free and responsible *because* everything we do *cannot* be explained in terms of our brain chemistry and genes. In contrast, the participants who failed the manipulation check on the PD scenario answer along the same lines as those who passed (64% *yes* to free will and 71% *yes* to responsibility).

As with most results in experimental philosophy, there are various ways to interpret the results of this survey. For instance, a libertarian interpretation might suggest that the ND scenario describes determinism in a way ordinary folk can internalize and they respond accordingly (with the 'proper' incompatibilist intuition), while the PD scenario describes determinism in a way that allows people to disregard it and assume that the Ertans (like us) have libertarian freedom over which thoughts, desires, and plans they have. Further testing might lend credence to such an interpretation. But, in conjunction with results from my prior collaborative work and from Nichols' experiments on intuitions about psychological determinism, I think the best interpretation of these results supports my hypothesis: most people do *not* regard psychological determinism to be a threat to free and responsible action but most people *do* regard reductionistic pictures that suggest 'bypassing' to be a threat to free and responsible action. If this hypothesis has any merit, it suggests that some incompatibilist arguments may derive some of their intuitive appeal by eliciting this 'derivative' intuition—that is, by describing determinism in a way that suggests such a reductionistic picture.

**4. Conflicting Intuitions**

I have offered some alternative interpretations of those results of Nichols' research that suggest the folk have libertarian conceptions of agency and incompatibilist intuitions about responsibility and determinism. Then I have used other parts of Nichols' research, along with my own experiment, to support my view that most people are compatibilists about *psychological* determinism and responsibility but incompatibilists about *reductionism* and responsibility, and that it does not seem that most people take determinism to entail reductionism. Well, what else would you expect from a compatibilist? I'm just trying to sell my own intuitions like any good philosopher (though I have at least supplemented my claims with some empirical work on pre-philosophical intuitions).

But I agree with Nichols that the empirical research that has been done on folk intuitions about freedom, determinism, and responsibility does suggest some interesting intuitional conflicts. I will close by briefly mentioning five such cases:

1) *The ability to do otherwise as unconditional vs. conditional.* I have already mentioned that people seem to express different intuitions about whether determinism precludes the ability to do otherwise and about whether responsibility requires an unconditional ability to do otherwise (i.e., holding fixed past conditions and the laws of nature). Alternatively, ordinary people may not have intuitions fine-grained enough to support either side of this debate.

2) *Free will requires alternatives vs. ownership.* Participants' explanations on the surveys suggest to me that they pick out two paradigms of free will that map roughly onto the paradigms philosophers develop into theories. Some people talk about the importance of alternatives for action (e.g., "they did not have a choice," "The Ertans could not do otherwise," and "Regardless of what happens in your past, you still make a choice of what you [are] going to do in the present"). Others talk about the agent acting on his or her own desires and beliefs (e.g., "the actions they do are still what they want to do (free will)," "It was <u>his</u> desire, <u>his</u> thoughts, <u>his</u> plan, hence his own free will," and "Because these thoughts are their own, so they act on their own thoughts").

3) *Free will as forward-looking (lack of) prediction vs. backwards-looking explanation*. It is possible that the intuitional conflicts about agency that Nichols focuses on derive largely from a temporal factor. When we consider the future and what we or another person might decide to do, we tend to focus on the various ways things could go, especially given the lack of information about influential events that have yet to occur. The uncertainty of prediction may prime intuitions about (indeterministically) open alternatives. But when we try to *explain* events that have already happened, we tend to want complete understanding and hence look for deterministic causal explanations. Our quest for certainty in explanation may prime intuitions about psychological determinism.

4) *Responsibility as forward-looking social control vs. backwards-looking retribution.* When it comes to intuitions about moral responsibility, intuitions and responses may vary depending on whether people are considering the forward-looking aspects of blame and punishment, aimed primarily at shaping future behavior, or the backwards-looking aspects, aimed primarily at harming the perpetrator to "balance the books." Some philosophers have suggested that the compatibilist conception of free will works to underpin the forward-looking notion of responsibility, while the libertarian conception is required to justify the retributive notion of responsibility.

5) *Praise vs. blame*. The previous point may also be one reason people's intuitions about the conditions required for praise (and reward) seem to differ from their intuitions about the conditions required for blame (and punishment). Susan Wolf (1990) suggests that we are willing to praise even if the person could not do otherwise while we are willing to blame only if the person could do otherwise. Some experimental results have suggested there may be some truth to this (see Nahmias *et al*. 2005). But it also seems that other conditions will prime people to blame but not to praise.

Now, all these conflicts and confusions might lead a skeptic about experimental philosophy to suggest that all this folk polling is a waste of time—it just tells us what we already knew, that these issues are complex and difficult and require the reflective consideration of well-trained philosophers. I agree that the issues are complex and require philosophical consideration, but I think that the reflective consideration of philosophers can be significantly informed by the results of surveying the folk. From such surveys we can get a better understanding not only of where the conflicts exist, but also of what drives people to have certain intuitions, what drives conflicting intuitions, and what intuitions people are more or less willing to give up in the face of conflicts (e.g., during 'reflective equilibrium'; see Nichols' discussion of this). If nothing else, the descriptive project of elucidating folk intuitions about freedom and responsibility offers interesting new material to drive the substantive and prescriptive projects forward.

**References**

Fischer, J. 1994. *The Metaphysics of Free Will*. Cambridge: Blackwell publishers.
Lewis, D. 1981. Are we free to break the laws? *Theoria* 47: 113-121.
Kane, R. 1999. Responsibility, luck, and chance: reflections on free will and indeterminism. *Journal of Philosophy* 96: 217-240.
Kim, J. 1998. *Mind in a Physical World.* Cambridge: MIT Press.
Nahmias, E., Morris, S., Nadelhoffer, T. & Turner, J. 2004. The phenomenology of free will. *Journal of Consciousness Studies* 11(7-8): 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*.
Nichols, S. 2004. The folk psychology of free will. *Mind and Language*, 19: 473-502.
Nichols, S. and Knobe, J. in prep. Moral responsibility and determinism: empirical investigations of folk intuitions.
Taylor, R. 1963. *Metaphysics*. Englewood Cliffs: Prentice-Hall.
Turner, J. and Nahmias, E. forthcoming. Are the folk agent-causationists? *Mind and Language*.
van Inwagen, P. 1975. The Incompatibility of Free Will and Determinism. In Watson, G. (ed.) *Free Will* (New York: Oxford University Press, 1982), 46-58.
van Inwagen, P. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.
Wolf, S. 1990. *Freedom within Reason*. New York: Oxford University Press.

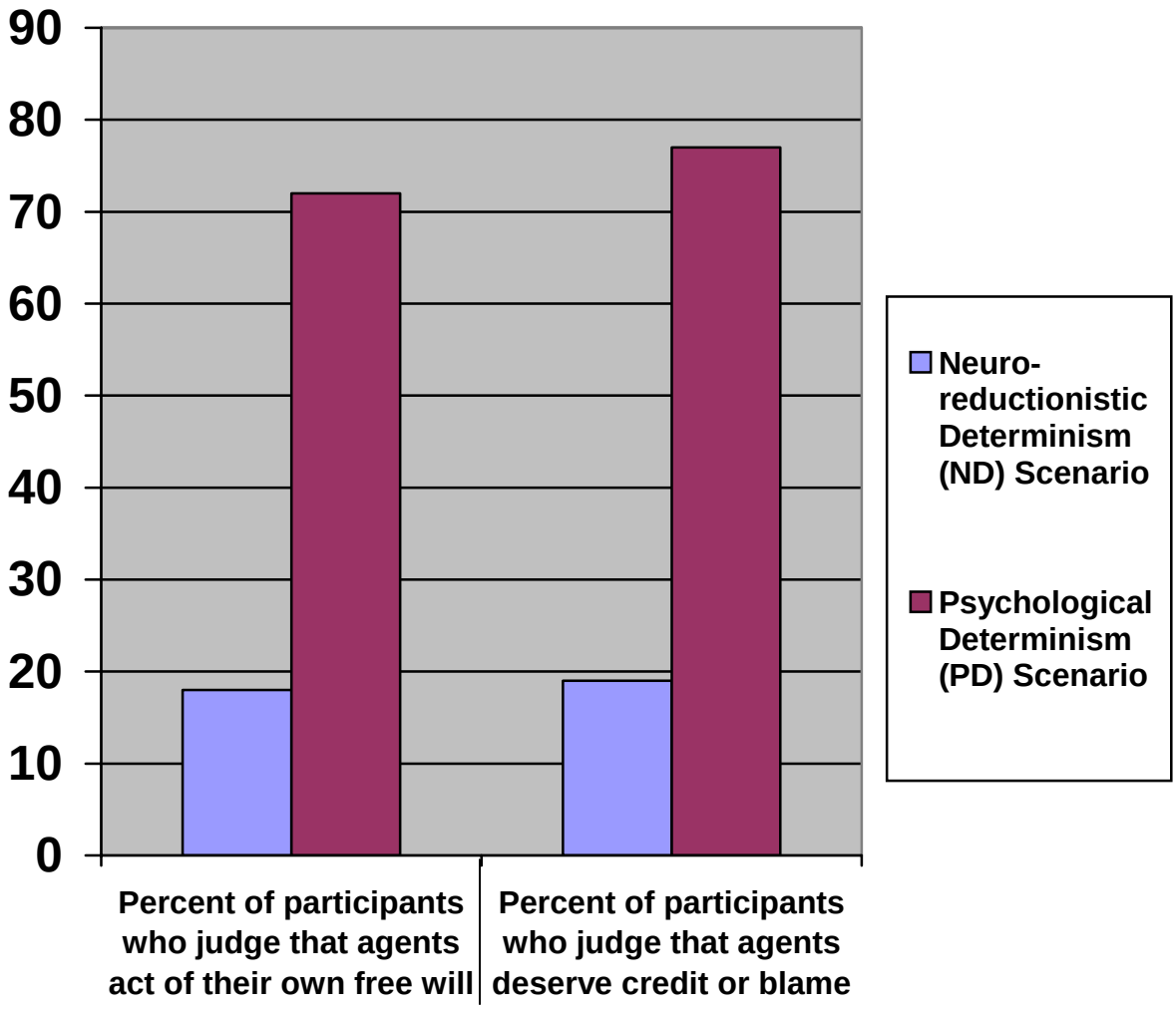


Figure 1: One set of participants (represented by blue bar) read a scenario describing agents whose actions are completely caused by their brain states, which were completely caused by their genes and environment. Another set of participants (represented by red bar) read a scenario describing agents whose actions are completely caused by their psychological states (desires, plans, etc.), which were completely caused by their genes and upbringing. Participants were then asked whether such agents (1) act of their own free will and (2) deserve credit or blame for their actions.