

Is Free Will Necessary for Moral Responsibility?: A Case for Rethinking Their Relationship and the Design of Experimental Studies in Moral Psychology

CARRIE FIGDOR AND MARK PHELAN

Abstract: Philosophical tradition has long held that free will is necessary for moral responsibility. We report experimental results that show that the folk do not think free will is necessary for moral responsibility. Our results also suggest that experimental investigation of the relationship is ill served by a focus on incompatibilism versus compatibilism. We propose an alternative framework for empirical moral psychology in which judgments of free will and moral responsibility can vary independently in response to many factors (including beliefs about determinism). We also suggest that, in response to some factors, the necessity relation may run from responsibility to free will.

1. Introduction: Freedom and Moral Responsibility

How are free will and moral responsibility related? As Kadri Vihvelin (2011) writes, within the philosophical literature on moral responsibility ‘it is ... widely agreed that the existence of free will is a necessary condition of the existence of moral responsibility.’ Michael McKenna (2009) seconds Vihvelin’s summary:

Free will is understood as a necessary condition of moral responsibility since it would seem unreasonable to say of a person that she deserves blame and punishment for her conduct if it turned out that she was not at any point in time in control of it.

Peter Van Inwagen (1983) also prominently defends this position, writing that:

... without free will there is no moral responsibility: if moral responsibility exists, then someone is morally responsible for something he has done or for something he has left undone; to be morally responsible for some act or failure to act is at least to be able to have acted otherwise, whatever else it may

CF would like to thank Yale Cognitive Science (Joshua Knobe and colleagues) for providing funding, development support and encouragement for the initial study, and Eddy Nahmias and Eric Mandelbaum for helpful comments. MP would like to thank Wesley Buckwalter, Eric Mandelbaum and Eddy Nahmias for helpful comments.

Address for correspondence: Carrie Figdor, Department of Philosophy, 260 EPB, Iowa City, IA 52245, USA.

Email: carrie-figdor@uiowa.edu

involve; to be able to have acted otherwise is to have free will. Therefore, if moral responsibility exists, someone has free will. Therefore, if no one has free will, moral responsibility does not exist (p. 162).

Finally, in *Freedom and Resentment*, P.F. Strawson acknowledges widespread agreement on this view. He distinguishes pessimists and optimists regarding the justifiability of practices of punishment (/reward) and blaming (/praising) under the assumption that determinism is true; pessimists think such practices would not be justified, optimists think they would. But he also notes that both sides agree that freedom is necessary for moral responsibility even though they differ on what 'freedom' means.

We are interested in examining this received, seemingly obvious, assumption that free will is necessary for moral responsibility, on any conception of freedom. We think this metaphysical question can be addressed in part by using the experimental method of contrast cases to answer a related question: do the folk agree with philosophical tradition that an agent must be free to be morally responsible? Of course, the folk could be wrong about what is needed for moral responsibility, whatever they might think. However, if what the folk believe departs from philosophical tradition, there is a case for reexamining the motivation for the traditional view, and a *prima facie* reason to hold that that view might be mistaken. The answer to the descriptive question therefore has important implications for the metaphysical one. We will set the stage for our answer to the descriptive question by summarizing previous experimental research on freedom and moral responsibility in the next section. We then present the results of our own experimental investigation starting in Section 3.

2. Prior Experimental Results: A Brief Literature Review

A natural place to look for insight into the descriptive question is the experimental philosophy of free will and moral responsibility. However, and in line with philosophical tradition, previous experimental work has focused on the compatibility or incompatibility of freedom *and* moral responsibility with determinism. Though this traditional debate is not our concern, it bears indirectly on the necessity question that is. Experimental philosophers have asked if people are natural incompatibilists—that is, do they pretheoretically hold that free will and moral responsibility are incompatible with causal determinism? This experimental question presupposes that what is true of free will vis-à-vis determinism is true of moral responsibility as well (at least in cases where there are no further reasons to undermine responsibility—this caveat will be taken for granted in what follows). This coupling of free will with moral responsibility is sanctioned by the received view illustrated above. Some philosophers have answered the compatibilism question in the affirmative, others in the negative. As experimental philosophers have pointed out, it is often unclear what evidence grounds traditional answers, and so in response they have employed experimental methods to obtain an empirically justified answer. But these methods have

not converged on an answer: some experimental results seem to favor incompatibilism, others compatibilism. This at least raises the question of whether experiments should continue to be designed along the lines of the traditional philosophical debate, in which the assumption that free will is necessary for moral responsibility is a fundamental principle. Here we briefly review some of the more influential of these studies in order to relate this literature to our concerns.

In their seminal experimental work on free will, Nahmias *et al.* (2005, 2006, 2007) challenge the naturalness of incompatibilism, arguing on the basis of original results that people are natural compatibilists about both free will and moral responsibility. In Nahmias *et al.* (2005), for example, they had people read 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 25 March 2150 AD, 20 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 PM on 26 January 2195. As always, the supercomputer's prediction is correct; Jeremy robs Fidelity Bank at 6 PM on 26 January 2195.

Participants in one condition were subsequently asked: 'Do you think that, when Jeremy robs the bank, he acts of his own free will?' In another condition, participants were asked: 'Do you think that, when Jeremy robs the bank, he is morally blameworthy for it?' Most participants in the first condition (76%) responded that Jeremy acts of his own free will. Most participants in the second condition (83%) responded that Jeremy is morally blameworthy. This (among other results) leads Nahmias *et al.* to conclude that, contra philosophical tradition, people are natural compatibilists about free will and moral responsibility. They further conclude that the consistency between the patterns of responses for these two questions 'accords well with philosophers' claims that judgments about free will are closely related to judgments about moral responsibility' (p. 568).

Nichols and Knobe (2007), on the other hand, contend that, contra Nahmias *et al.*, people are natural incompatibilists when it comes to moral responsibility, but that they are drawn towards compatibilism in affect-laden contexts:

[M]ost people (at least in our culture) really do hold incompatibilist theories of moral responsibility, and these theories can easily be brought out in the kinds of philosophical discussions that arise, e.g. in university seminars. It's just that, in addition to these theories of moral responsibility, people also have immediate affective reactions to stories about immoral behaviors (p. 668).

For instance, in one of their studies participants read about a causally determined universe ‘in which everything that happens is completely caused by whatever happened before it’ (p. 669). One group of participants—those in the ‘concrete condition’—was then asked whether a particular agent, Bill, who lives in this causally determined universe, could be fully morally responsible for killing his wife and children to be with his secretary. Another group of participants—those in the ‘abstract condition’—was asked a generic question of whether it is possible for a person living in the causally determined universe to be fully morally responsible for his or her actions. Nichols and Knobe found that while most participants in the concrete condition (72%) responded as compatibilists, holding Bill fully morally responsible for killing his wife and children, most participants in the abstract condition (84%) responded as incompatibilists, denying that it is possible for a denizen of a deterministic universe to be fully morally responsible for their actions.

Notably, Nichols and Knobe set aside questions concerning compatibilism or incompatibilism of determinism and *free will*. So while their conclusions are in tension with Nahmias *et al.*’s about moral responsibility—in the sense that they disagree that the folk are natural compatibilists about responsibility—they do not address Nahmias *et al.*’s assessment that judgments about free will and moral responsibility vary in tandem. In effect, by targeting only the relation of determinism and moral responsibility, their study divides the incompatibilism question into two: whether judgments of free will are compatibilist or incompatibilist, and whether judgments of moral responsibility are compatibilist or incompatibilist.

Sinnott-Armstrong (2008) challenges Nichols and Knobe’s conclusion that the presence or absence of affective reactions is the primary explanatory factor for participants’ conflicting assessments of moral responsibility in the concrete and abstract cases. Sinnott-Armstrong argues that concreteness or abstractness in themselves can explain a large degree of the variance in people’s judgments. Nichols and Knobe had considered this possibility and attempted to rule it out experimentally. They asked college-aged participants about two concrete cases that differed in terms of emotional affect. In one case, Mark cheats on his taxes. In the other, Bill stalks and rapes a stranger. They found that participants were almost three times more likely to hold Bill responsible in a deterministic universe than Mark. Sinnott-Armstrong counters that these findings do not rule out concreteness as an explanatory factor, nor do they confirm the affective hypothesis, because the Bill case is more concrete than the Mark case as well as more affectively charged. There are many ways to cheat on your taxes, just as there are many forms of rape (e.g. date, marital). By including the information that Bill *stalks* his victim, the Bill case rules out various forms of rape. The Mark case, on the other hand, does not narrow the field of possible tax cheating scams. Moreover, the college students who participated in the study are presumably less likely to have a concrete conception of the details of filing taxes and more likely to have concrete conceptions of sex acts.

Sinnott-Armstrong argues instead that the results support a dual systems account of intuition in general. On his view, some of our intuitive judgments are the

deliverances of one representational system related to episodic memory that is engaged when we are dealing with concrete episodes. Others are the deliverances of a separate representational system related to semantic memory that is engaged when we are dealing with more abstract instances and principles. This position suggests that compatibilist and incompatibilist intuitions are equally natural, since they follow from different systems of judgment. However, Sinnott-Armstrong, like Nichols and Knobe, only discusses incompatibilism or compatibilism regarding moral responsibility. Thus, his paper also does not speak to the relationship between such judgments and judgments of free will.

Roskies and Nichols (2008) seem to lend support to Sinnott-Armstrong's claim that concreteness makes an important difference to judgments of compatibilism regarding moral responsibility. In their study, Roskies and Nichols had one group of participants read a story about an alternate universe in which many 'eminent scientists have become convinced [of determinism,] that ... every decision a person makes is completely caused by what happened before the decision.' Another group of participants read a story about our own universe—a story in which many of our own eminent scientists have reached the same deterministic conclusion about us. Nichols and Roskies then elicited participants' levels of agreement with a series of abstract statements about the possibility of moral responsibility, blame, and free will in the universe they had read about, on the assumption that its scientists were right about determinism.

They found that participants tended to agree that it is impossible for people in the deterministic alternate universe to be either fully morally responsible or to make truly free choices. However, participants disagreed significantly more with statements denying moral responsibility and free will to people in our own (presumed) deterministic universe. These results accord with Sinnott-Armstrong's abstract/concrete hypothesis in that we can be expected to have very concrete ideas of our own world but not about an alternate universe. Notably, their results also appear to lend support to the philosophical tradition that free will is necessary for moral responsibility, since both kinds of judgments were affected in the same way by manipulations of concreteness.

While we have only summarized here some of the most influential papers in the large experimental literature on the relation between determinism and free will and responsibility, one thing that seems clear is that this research doesn't definitively answer the question of whether people are natural compatibilists or incompatibilists. There is clearly some contextual variance in people's assessments of the degree to which free will and moral responsibility are compatible with determinism. Moreover, in the absence of a theory of intuitive judgment, we don't know which if any of the variant assessments to privilege over others. In addition, some have raised the worry that the studies aren't really getting at the compatibilism issue at all because participants are misinterpreting the concept of determinism involved (Sommers, 2010) or conflating determinism with reductive mechanism (Nahmias *et al.*, 2007).

For our purposes, the more important issue is that these studies (and the extant research in general) are not focused on the question of whether free will is required

for moral responsibility. Experimental studies typically have been designed to explore how the determinism factor affects assessments of freedom and responsibility. In light of the diverse results of these studies, new factors have been proposed to explain how freedom and responsibility judgments may be modulated given determinism or indeterminism, and the results are interpreted as either supporting compatibilism or incompatibilism. The necessity question is overlooked, along with the possibility that the results may be showing that manipulating the determinism factor affects assessments of freedom and moral responsibility asymmetrically. In any case, as traditionally understood the necessity of freedom for moral responsibility would imply that for any manipulation whatsoever an agent should not be judged morally responsible without also being judged free (or that attributions of moral responsibility would positively correlate with attributions of freedom). We cannot tell whether this necessary connection holds without manipulating a variety of factors to see whether there are conditions in which participants judge that an agent was morally responsible while lacking free will.

We therefore took an atypical experimental tack. The studies just discussed look at a limited set of factors that might affect judgments of freedom or moral responsibility—determinism, abstractness, emotional valence—but do not investigate whether judgments of freedom and moral responsibility might constitute importantly distinct assessments, which may be influenced in different ways by different factors. The latter approach calls for a framework for study design that is compatible with distinct connections, given that the folk may not be of one mind on this issue (as with much else). In a series of studies, we systematically varied factors related to the ultimate scientific explanation of human behavior (psychology versus neuroscience) and the predictability of human action (eventually predictable versus remaining unpredictable). We then investigated how people would assess freedom and moral responsibility under each variation. Are there really *no* instances, in these conditions at least, in which people judge that someone lacks freedom though they have moral responsibility?

3. First Study: Predictability and Explanatory Domain

3.1 Methods

We initially sought to investigate how predictability, not determinism, might affect judgments of free will and moral responsibility. That is, we questioned whether the compatibilism versus incompatibilism framework was adequate for interpreting the emerging research record, and so sought to detach free will and responsibility judgments from the question of determinism entirely. We did not, at the outset, also question or explicitly test the received view of their necessary relation. Historically the relationship between freedom and responsibility, on the one hand, and predictability has been of immense interest. Medieval philosophers focused on the relation of divine foreknowledge to freedom and moral responsibility. But recent theorists have also discussed foreknowledge in connection with Frankfurt's (1969)

well-known challenges to the Principle of Alternative Possibilities (see Zagzebski, 1991, for discussion). We wanted to investigate whether people would attribute less freedom and responsibility to agents whose actions were fully scientifically predictable. We predicted that they would.

We varied the predictability factor independently of scientific explanation, since sciences concerned with explaining and predicting decisions and behavior are asymmetrical with respect to their perceived threat to free will. Neuroscientific explanation is often perceived as a *prima facie* threat to free will, whereas belief-desire explanation has coexisted with free will attributions for centuries. We reasoned that folk judgments of freedom and responsibility would be affected by the manner of explanation of our actions and decisions. Thus, we varied explanatory domain (neuroscientific, belief-desire) across vignettes as well as predictability. Our prediction was that agents would be ascribed less freedom and less moral responsibility if their decisions and actions were explainable in neuroscientific terms rather than in belief-desire terms.¹ Finally, we manipulated the concreteness of our vignettes by varying whether they occurred on Earth or Erta (as in Nahmias *et al.*, 2007, and Roskies and Nichols, 2008). We expected to replicate previous findings in which moral responsibility and freedom are attributed at higher levels in concrete scenarios.

This yielded a 2 (Predictability: predictable, not predictable) x 2 (Domain of Explanation: neuroscience, belief-desire) x 2 (Concreteness: Earth, Erta) study in which subjects read scenarios in which either neuroscientists or psychologists will soon be able (or will never be able) to predict decisions and actions with high accuracy, based on increased knowledge of their causes, described in their respective causal-explanatory kind terms (e.g. biochemical processes versus desires). For example, this is our vignette for the predictable*neuroscience*Earth condition (with variations for the 'belief-desire' condition bracketed and bolded):

Most neuroscientists **[psychologists]** are convinced that scientists will soon figure out exactly how our decisions and actions are caused. They think that whenever we are trying to decide what to do, the decision we end up making is completely caused by the specific chemical reactions and neural processes **[thoughts, desires and plans]** occurring in our brains **[minds]**. The neuroscientists **[psychologists]** also hold that these chemical reactions and neural processes **[thoughts, desires and plans]** are completely caused by our current situation and our brain **[psychological]** processes that occurred earlier in our lives.

Moreover, they agree that this knowledge of chemical and neural **[psychological]** causes will enable scientists to predict our decisions and actions with much

¹ Nahmias *et al.* (2007) combined a similar domain-of-explanation manipulation with a determinism/indeterminism manipulation that was designed to test their hypothesis that mechanism, not determinism, drives free will and moral responsibility judgments.

greater accuracy and precision than they can today. For instance, they will be able to predict whether you will decide to eat a banana today and when you will eat it if you do. They believe that the reason why our decisions and actions are not highly predictable now is because we lack this knowledge. Once we have this chemical and neural **[psychological]** information, predictability will follow.

Here is the vignette for the predictable* neuroscience* Erta condition (with variations for the 'non-predictable' condition bracketed and bold):

On Erta, the landscape and life are very similar to Earth, and there are advanced life forms called Ertans who look, talk, and behave very much like we do. In particular, they ascribe beliefs, wants, intentions and other psychological states to each other, and they rely on these psychological states to explain each others' behavior, just as we do.

However, unlike on Earth, Ertan neuroscientists believe that they will soon figure out exactly how Ertans' decisions and actions are caused. They believe they will be able to describe the specific chemical reactions and neural processes occurring in Ertans' brains when they make decisions. They also believe that this information will enable them to predict anything any Ertan will do with extremely high accuracy and precision. **[But they do not believe that this information will enable them to predict anything any Ertan will do with any more accuracy or precision than they can now.]** For example, they will **[not]** be able to predict whether an Ertan will decide to eat a banana on a particular day or when he or she will eat it if he or she does. By way of comparison, **[even]** if human neuroscientists were to learn this much about human brains, they would also be able to predict your behavior with much greater accuracy and precision than they can now **[they would still not be able to predict your behavior with any more accuracy or precision than they can now]**.

In all vignettes, we used the case of deciding to eat a banana as a concrete but non-moral and non-affect-laden example of an action.

At the beginning of our study, participants responded to two brief comprehension questions, aimed at testing whether they read carefully. After responding to these comprehension questions, participants read the vignette to which they had been assigned. Participants were then asked to indicate their level of agreement with five statements pertaining to free will and moral responsibility. For example, participants in the Earth*belief-desire conditions were asked to indicate their agreement with the following statement (among four others): 'If the psychologists are right, we are able to make decisions of our own free will.' Participants indicated their agreement with each statement on a seven-point Likert-scale anchored at 'strongly disagree' (1) and 'strongly agree' (7). In addition to free will, participants were asked to indicate

their agreement with statements affirming moral responsibility, blameworthiness, praiseworthiness, and whether our (or the Ertans') actions are up to us (or them).²

3.2 Results and Discussion

3.2.1 Main Effects. We will discuss main effects for our three manipulations (predictability, concreteness, and domain of explanation) in this sub-section, before discussing interaction effects in 3.2.2. Participants who did not pass either of the comprehension checks at the start of the study were not included in our analyses in this or in any of the subsequent studies described below. Results for the remaining 147 participants were subjected to statistical analysis. Contra our prediction above, we found no significant difference related to predictability.³ However, corresponding to our prediction above, there was an overall effect for concreteness, and significant effects of concreteness for *morally responsible*, *blameworthiness*, and *praiseworthiness*, although not for *up to us* or *free will* (see Figure 1 for a comparison of *free will* and *moral responsibility*).⁴

These effects suggest that judgments of moral responsibility, praiseworthiness and blameworthiness tend to hang together, while judgments for the freedom and up-to-us prompts cluster. Because freedom and moral responsibility are the

² Participants in this and the follow-up studies discussed below were recruited and tested using commercially available online platforms (Qualtrics and Amazon Mechanical Turk). In this initial study, there were 229 participants. Two-hundred thirteen supplied gender data, of which 54% identified themselves as male and 46% as female. Twenty-five percent of participants were aged 18–25, another 40% were 26 to 35, and 95% of participants were under 65. Fifty percent of participants had not attained a college degree, and 86% had not gone beyond a four-year college degree. Participants were paid between \$0.25 and \$0.50 for their participation. Participants were located in the United States and were prohibited from taking part in more than one study.

³ A 2 (Predictability) $\times$ 2 (Domain of Explanation) $\times$ 2 (Concreteness) Multivariate ANOVA was used to compare the influence that predictability had on up to us, free will, moral responsibility, blameworthy, and praiseworthy judgments. The multivariate result was not significant for predictability: Pillai's Trace = 0.019, $F(5, 133) = 0.518$, $p = 0.762$. T-tests were subsequently used to assess the effect of predictability on participants' agreement with each ascription statement (e.g. on participants' agreement with the free will ascription statement). Predictability made no significant difference in participants' agreement with any of these ascription statements. We did find a small main effect for predictability for judgments of moral responsibility, but not for free will, when all 229 participants were included in the analysis. Univariate F tests showed there were significant differences between attributions of *Morally Responsible* $F(1, 216) = 4.724$, $MSE = 2.653$, $p = 0.031$, and *Blameworthy* $F(1, 216) = 4.874$, $MSE = 2.749$, $p = 0.028$, and a marginally significant effect for *Praiseworthy* $F(1, 216) = 3.764$, $MSE = 2.729$, $p = 0.054$, while no significant difference with respect to prediction was detected for *free will* or *up to us* (though they were approaching for *Up to us* $F(1, 216) = 2.818$, $MSE = 2.977$, $p < 0.095$).

⁴ The multivariate result was significant for concreteness (Pillai's Trace = 0.108, $F(5, 133) = 3.204$, $p = 0.009$). With respect to concreteness, T-tests showed significant differences between attributions of *Morally Responsible* $t(145) = 3.069$, $p = 0.003$, *Blameworthy* $t(145) = 2.64$, $p = 0.009$, and *Praiseworthy* $t(145) = 2.711$, $p = 0.008$. No significant difference with respect to concreteness was detected for *Up to us* or *Free will*.

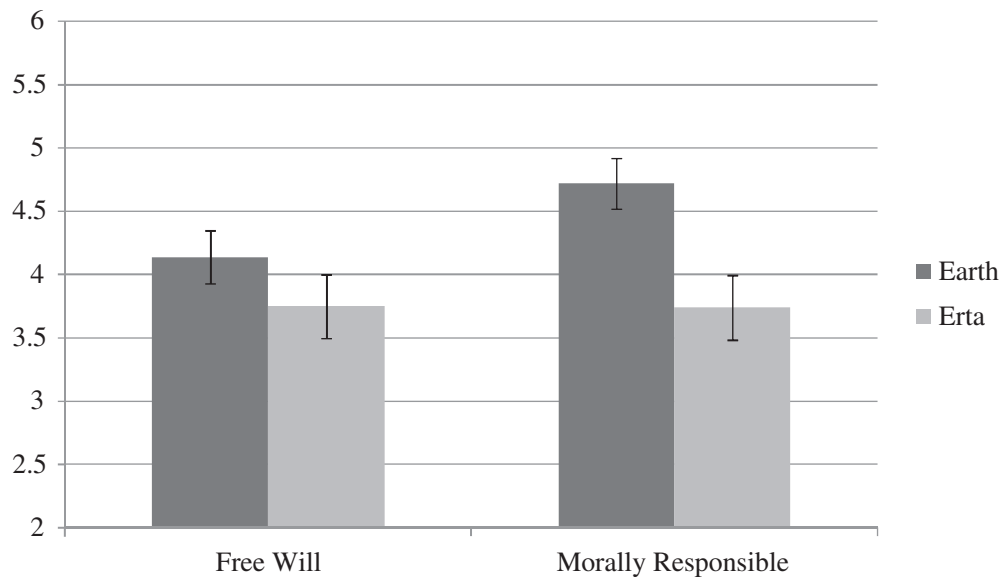


Figure 1 Mean agreement with free will and moral responsibility attributions for the concreteness manipulation.

concepts we are primarily interested in, we will focus on results for these two judgments in what follows.

As noted, we found one of our predicted significant effects for moral responsibility judgments: people tended to agree more with attributions of moral responsibility in concrete scenarios. This result for concreteness replicates Nahmias *et al.*'s (2007) and Roskies and Nichols' (2008) findings that varying concreteness significantly affects moral responsibility (and blameworthiness) judgments. However, we were surprised that this effect appeared for moral responsibility assessments but not for free will assessments (although there was a trend in the appropriate direction). The necessity hypothesis suggests there should be no difference.⁵ We will return to this tension with the necessity hypothesis below.

The largest main effects we found were for domain of explanation. Agreement with attributions for each of our dependent variables (up to us, free will, morally responsible, blameworthiness, and praiseworthiness) were lower when neuroscientists were doing the predicting (using neurochemical kinds) than when psychologists were (using psychological kinds).⁶ (See Figure 2 for representative results for free

⁵ Feltz and Cokely (2009), among others, have found that individual differences affect judgments of free will and moral responsibility. We agree that an individual difference effect is also likely in folk assessments of the free will/moral responsibility relationship, but think this hypothesis would best be addressed by research focused on that question.

⁶ The multivariate result was significant for domain of explanation (Pillai's Trace = 0.216, $F(5, 133) = 7.344$, $p < 0.001$). For each dependent variable (up to us, free will, morally responsible, blameworthy, and praiseworthy), T-tests showed a significant difference between attributions (each $p < 0.001$) with respect to the domain of explanation.

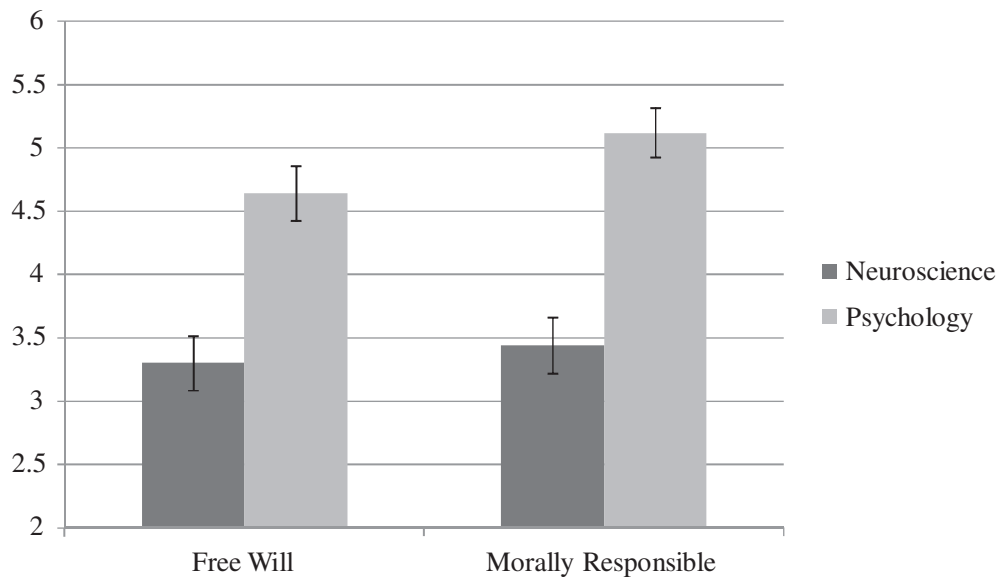


Figure 2 Mean agreement with free will and moral responsibility attributions for the domain of explanation manipulation.

will and moral responsibility.) This is consistent with Nahmias *et al.*'s (2007) result in which neuroscientific explanation led to less agreement with these five variables (compared to psychological explanation) in both deterministic and indeterministic scenarios.

For each of the five variables, mean attributions were below the midline (of 4) if actions and decisions were characterized as being fully explicable in terms of chemical reactions and neural processes. They were all well above midline if actions and decisions were characterized as being fully explicable in terms of beliefs and desires. On one explanation of this domain of explanation effect, the real threat to free will and moral responsibility is neuroscientific explanation, which may seem to foreclose the possibility of a non-material mind. Alternatively, this effect might be explained in terms of an authority effect, whereby neuroscientists are considered real scientists (or neuroscience a real science) and psychologists are not (or psychology is not a science). However, we set aside this question for future work in order to focus on the necessity question.

3.2.2 Interaction Effects. We turn now to interaction effects and the relationship between freedom and moral responsibility. Though we sometimes found differences in main effects for freedom and moral responsibility, the judgments were generally of a piece. Responses to free will and moral responsibility prompts trended in the same direction, and tended to fall on the same side of the midline within conditions. In light of this general agreement, two interaction effects are of particular interest in assessing whether free will is required for moral responsibility. There was a significant interaction effect for predictability*domain of explanation on free will

judgments, but not on moral responsibility judgments.⁷ Levels of agreement with the free will statement were higher in all belief-desire conditions than in the neuroscience conditions, and were much higher in the low-predictability belief-desire conditions than in all other conditions. There was no such effect for moral responsibility assessments, although there was a non-significant trend in the same direction as with free will. Thus, this finding, while interesting, does not contradict the claim that freedom is required for moral responsibility.

However, a more surprising result, and one that tells against the received view that freedom is required for moral responsibility, involves an interaction effect between domain of explanation*concreteness for moral responsibility judgments, but not free will judgments.⁸ As shown in Figure 3, moral responsibility judgments drop precipitously when comparing the neuroscience*Earth scenarios to the neuroscience*Erta scenarios.⁹

Of particular importance to our project of investigating whether free will is necessary for moral responsibility is the low level of free will judgments in the neuroscience*Earth vignettes, compared to moral responsibility judgments. As shown in Figure 4 (a close-up of the right half of Figure 3), these judgments straddled the midline, suggesting that participants tended to agree that people lacked free will in this vignette, though they agreed that they were still morally responsible.

This result is directly in tension with the hypothesis that free will is necessary for moral responsibility, and inspired the four studies that follow.¹⁰

⁷ An ANOVA showed a significant difference between attributions of *free will* with respect to predictability*domain of explanation: $F(5, 137) = 4.465$, $p = 0.036$; but not for attributions of *moral responsibility*: $F(5, 137) = 3.304$, $p = 0.071$.

⁸ An ANOVA showed a significant difference between attributions of *moral responsibility* with respect to domain of explanation*concreteness: $F(5, 137) = 8.047$, $p = 0.005$; but not for attributions of *free will*: $F(5, 137) = 0.796$, $p = 0.374$.

⁹ A difference in folk intuitions in abstract and concrete conditions (in our study and others) does raise the question as to which of these conditions prompts the true folk view. In this study in particular, which condition prompts participants' theoretically relevant intuitions about moral responsibility: neuroscience*Earth (our concrete condition) or neuroscience*Erta (our abstract condition)? Nichols and Knobe (2007) consider several frameworks for moral judgment that give different answers to this question. If judgments of moral responsibility are biased by affect in concrete conditions (the 'performance error' theory), then abstract conditions reflect true folk judgments; if affect is a core component of moral judgment (the 'affective competence' theory), then concrete conditions do. After considering these and two other possibilities, they ultimately leave this question open for further study. Without settling the issue either, we consider that our concrete conditions reflect participants' true judgments on the basis that it is reasonable to expect the real world to be more vivid (literally concrete) to the folk than a hypothetical planet. Moreover, since our scenarios involve benign actions, such as eating a banana, they are unlikely to engender potentially distorting affective responses, which may be aroused when participants read about an agent killing his wife or raping a stranger.

¹⁰ Nahmias *et al.* (2007) also found a difference in judgments of moral responsibility when testing whether a difference in goodness or badness of a concretely described action would mediate the effect of neuroscientific explanation. MR responses were significantly higher between Bad

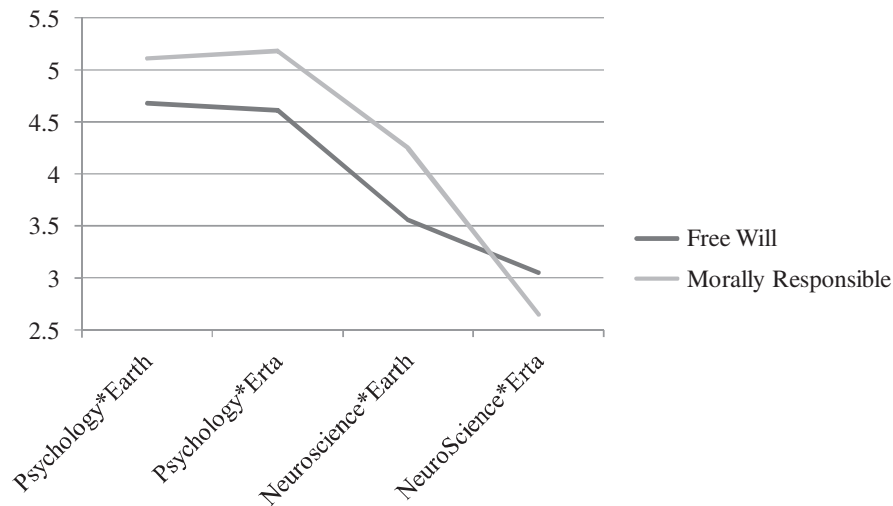


Figure 3 Mean agreement with free will and moral responsibility attributions for domain of explanation*concreteness manipulation.

4. Second Study: Forced Choice for Free Will and Moral Responsibility

4.1 Methods

In study 1, we found a domain of explanation*concreteness interaction effect for moral responsibility judgments, but not free will judgments. Of particular interest in assessing the hypothesis that freedom is required for moral responsibility was the fact that participants tended to agree that people could be morally responsible in situations in which they tended to agree they did not have free will, namely, in neuroscience*Earth scenarios. Given this asymmetry, we decided to investigate whether people would deny free will to humans if neuroscience were capable of explaining our behavior, while still attributing moral responsibility. In study 2, we used a forced choice response task in which participants read the neuroscience*concrete*not predictable scenario from study 1:

Most neuroscientists are convinced that scientists will soon figure out exactly how our decisions and actions are caused. They think that whenever we are trying to decide what to do, the decision we end up making is completely caused by the specific chemical reactions and neural processes occurring in our brains. The neuroscientists also hold that these chemical reactions and neural processes themselves are completely caused by our current situation and our brain processes that occurred earlier in our lives.

However, they agree that this knowledge of chemical and neural causes will not enable scientists to predict our decisions and actions with any greater accuracy

versus Good scenarios, whereas no corresponding significant difference was found for free will and up-to-us.

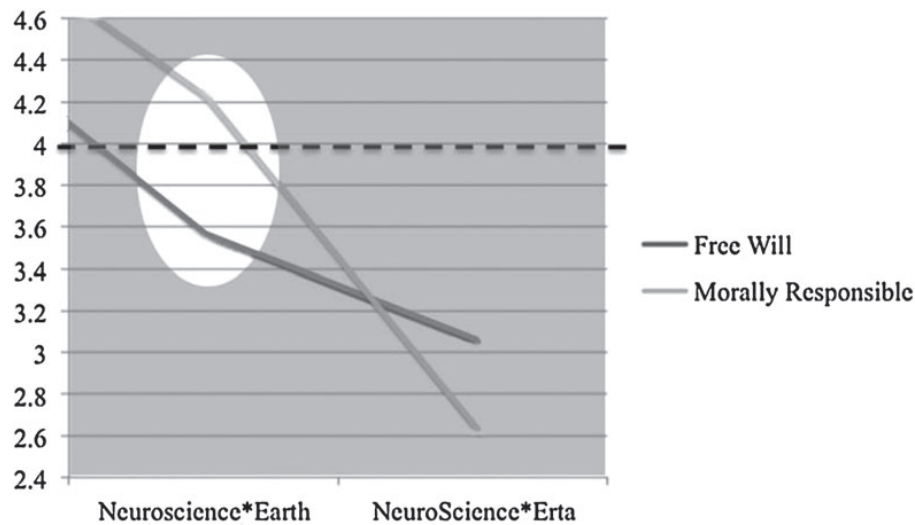


Figure 4 Mean agreement with free will and moral responsibility attributions for Neuroscience* Earth condition: mean moral responsibility assessment: 4.25; mean free will assessment: 3.56. (Dashed line marks midpoint.)

or precision than they can today. For instance, they will never be able to predict whether you will decide to eat a banana today or when you will eat it if you do. They believe that the reason why our decisions and actions will not be more predictable is because there are so many relevant factors involved in bringing them about. So even when we have this chemical and neural information, predictability will not follow.

Participants in a *free will condition* were asked, 'If the neuroscientists are right, are we able to make decisions of our own free will?' A different set of participants, in a *moral responsibility condition*, were asked, 'If the neuroscientists are right, should we be held morally responsible for our actions?' Subjects in both groups had to choose a yes or no response.¹¹

4.2 Results and Discussion

Once again, we excluded participants who failed either of our comprehension questions from our analysis. As shown in Figure 5, the remaining 128 participants were

¹¹ Participants were located in the United States. One-hundred forty-eight participants took part in this study. Fifty-eight percent of participants identified themselves as male and 42% as female. (All participants responded to the gender question.) Twenty-four percent of participants were aged 18–24, another 50% were 25 to 34, and 99% of participants were under 65. Forty-nine percent of participants had not attained a four-year college degree, and 91% had not gone beyond such a degree in their education. Participants were paid between \$0.20 and \$0.40 for their participation. Participants were prohibited from participating more than once.

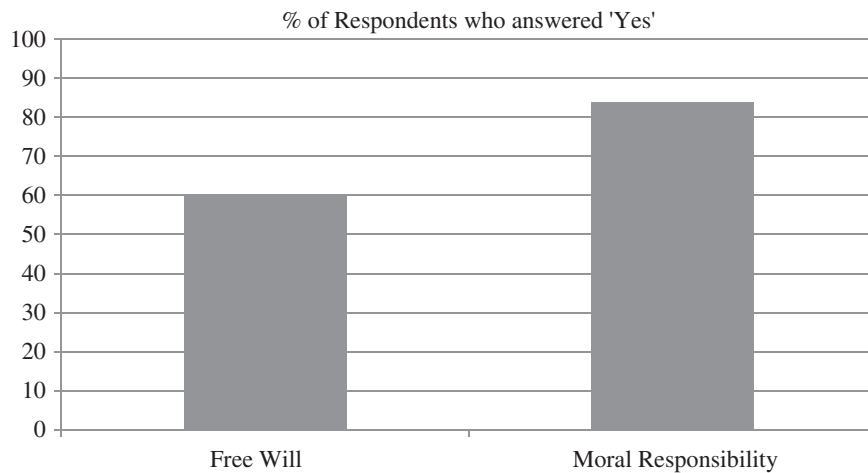


Figure 5 Percent agreement with free will and moral responsibility attributions in the between-participants, forced-choice neuroscience*concrete*not predictable study.

significantly more likely to say that we should be held morally responsible than that we have free will in a situation in which actual neuroscientists are able to fully explain our behavior but not to fully predict it.¹²

This finding suggests that for many ordinary English speakers considering a reasonable scenario in the actual world, free will is not required for moral responsibility (%'s are 60 and 84 respectively). It would be inappropriate to interpret this result as evidence for a compatibilist or a soft-compatibilist (non-libertarian) view of freedom. We did not probe possible differences in the conceptions of freedom participants might have had. Thus, the result is consistent with both libertarian and soft-compatibilist conceptions of freedom (or any other conception). Further research might attempt to determine and distinguish among different conceptions to see if the overall effect can be attributed to one particular conception rather than another and whether a subgroup of classical compatibilists can be identified. This subgroup would include those who think that freedom means lack of coercion, that we are both free and morally responsible, and that determinism is true.

5. Third Study: Within-Participants Forced Choice for Free Will and Moral Responsibility

5.1 Methods

Study 2 found between-subject differences in free will and moral responsibility judgments, but perhaps this is because asking about moral responsibility triggers associations (e.g. with the harm caused by moral violations) that are not triggered

¹² A chi-square test revealed a significant difference for free will and moral responsibility judgments: $\chi^2(1, N = 128) = 8.879, p = 0.003$.

when asking only about free will. If people really think moral responsibility doesn't require free will, we might expect to find this effect even when both questions are asked together.¹³ Thus, in study 3 we adopted a within-participants study design. We presented participants with the same vignette as in study 2, but this time, after reading the story, all participants answered both the free will and the moral responsibility questions from study 2.¹⁴ The order of these questions was randomized.

5.2 Results and Discussion

As in previous studies, participants who failed either of two comprehension controls appearing at the beginning of this study were excluded from our analysis. The results for the remaining 108 participants were subjected to McNemar's test. As in study 2, participants responded to these questions in significantly different ways. As shown in Figure 6, even when they responded to both questions, participants were significantly more likely to say that we should be held morally responsible than that we have free will in a situation in which neuroscientists are able to fully explain our behavior but not to fully predict it.¹⁵

Finally, we examined the effect of order on participant responses. Interestingly, when participants are split between those who received the free will question first and those who received the moral responsibility question first, we see that most of the variance in participants' responses to the questions is explained by order. Those participants who received the free will question first were significantly more likely to offer 'yes' responses to the moral responsibility question than to the free will question, while there was no difference in 'yes' responses for the two questions when the moral responsibility question appeared first.¹⁶ This might be due to associations

¹³ There may also be reason to think we would not find this result when both questions are asked together. For example, a consistency effect might lead some proportion of participants to respond in the same way to both questions, especially since some participants might respond quickly and less reflectively to online study probes.

¹⁴ Participants were located in the United States. One-hundred thirteen participants took part in this study. Fifty-five percent of participants identified themselves as male and 45% as female. Twenty-three percent of participants were aged 18–24, another 55% were 25 to 34, and 100% of participants were under 65. (All participants responded to the gender and age questions.) Fifty-six percent of the 110 participants who responded to an education question had not attained a four-year college degree, and 93% had not gone beyond such a degree in their education. Participants were paid \$0.40 for their participation. Participants were prohibited from participating more than once in this study.

¹⁵ An exact McNemar's test determined that there was a statistically significant difference in the proportion of 'yes' responses to the free will and moral responsibility questions, $p = 0.036$.

¹⁶ An exact McNemar's test determined that there was a statistically significant difference in the proportion of 'yes' responses to the free will and moral responsibility questions for those 52 participants who responded to the free will question first, $p = 0.007$. Twenty-nine responded 'yes' to the free will question, 40 responded 'yes' to the moral responsibility question. On the other hand, an exact McNemar's test found no statistically significant difference in the proportion of 'yes' responses to the free will and moral responsibility questions for those 56 participants who

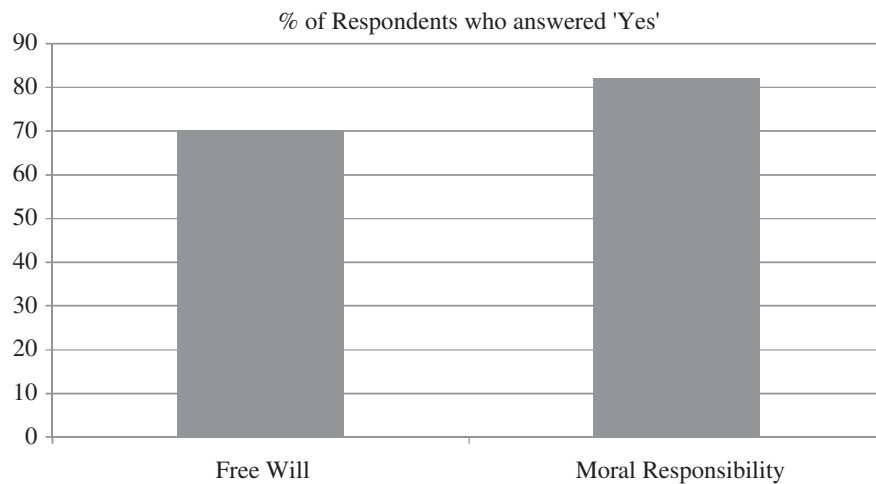


Figure 6 Percent agreement with free will and moral responsibility attributions in the 1st within-participants, forced-choice neuroscience*concrete*not predictable study.

with moral responsibility: if participants associate responsibility with possible harms, they may be more likely to also want to punish the responsible party, and ascribing them free will would facilitate possible punishment. On the other hand, if moral responsibility implies free will, this result shows that participants asked the free will question first *deny* the *modus tollens* (i.e. if not free, then not morally responsible), while participants asked about moral responsibility first *affirm* the *modus ponens* (i.e. if morally responsible, then free). In that case, the tradition must still explain why when someone is judged to lack free will, that person is not also judged to lack moral responsibility, contra van Inwagen's assertion above. We would not expect this pattern of responses if moral responsibility were thought to require freedom.

6. Fourth and Fifth Studies: Refinements of the Within-Participants Study

6.1 Methods for Study Four

A worry for the foregoing discussion emerges from careful consideration of the materials used in previous studies. To wit, the free will and moral responsibility questions are not perfectly balanced. The free will question is '... *are we* able to make decisions of our own free will'; the moral responsibility question, on the other hand, is '... *should we* be held morally responsible' (italics added here to emphasize the difference). But 'should' claims admit of various interpretations. After all, one might argue that we should treat people as though they are morally responsible, even if they really aren't. Thus, the moral responsibility question should be

responded to the moral responsibility question first, $p = 1.0$. Forty-one responded 'yes' to the free will question, 42 responded 'yes' to the moral responsibility question.

revised, to ask simply whether people *are* morally responsible. Study 4 used the same design and materials as study three, but made this one change in the moral responsibility question. So, participants in this experiment read the same vignette as in studies two and three, and responded to counter-balanced free will and moral responsibility questions as in study three, but, in this study, the moral responsibility question asked, 'If the neuroscientists are right, are we morally responsible for our actions?'¹⁷

6.2 Results and Discussion of Study 4

As in previous studies, participants who failed either of two comprehension controls appearing at the beginning of this study were excluded from our analysis. Results for the remaining 104 participants were subjected to McNemar's test. Unlike in studies 2 and 3, participants in study 4 did not respond to these questions in significantly different ways.¹⁸ For this study, only one result was approaching significance. For those participants who received the free will question first, the difference between assessments of free will and moral responsibility were significant at the 90% confidence level, but not at 95%.¹⁹ There was no difference in 'yes' responses for the two questions when the moral responsibility question appeared first.²⁰

The lack of a difference in this study may be partly explained by hypothesizing that the presence of the 'should' language in previous iterations of the moral responsibility question may have helped participants conceptualize moral violations or possible harms when they made their responsibility assessments. If this is correct, then using an example of a mildly immoral act in the vignette should mitigate in favor of a significant difference between the free will and moral responsibility questions. In study 5, we thus changed the act involved in these vignettes.

¹⁷ Participants were located in the United States. One-hundred eleven participants took part in this study. Sixty-four percent of participants identified themselves as male and 36% as female. Thirty-one percent of participants were aged 18–24, another 44% were 25 to 34, and 100% of participants were under 65. (All participants responded to the gender and age questions.) Sixty-two percent of the 110 participants who responded to our education question had not attained a four-year college degree, and 91% had not gone beyond such a degree in their education. Participants were paid \$0.40 for their participation. Participants were prohibited from participating more than once in this study.

¹⁸ An exact McNemar's test returned no statistically significant difference in the proportion of 'yes' responses to the free will (61%) and moral responsibility (65%) questions, $p = 0.332$.

¹⁹ An exact McNemar's test revealed a marginally significant difference in the proportion of 'yes' responses to the free will (56%) and moral responsibility (68%) questions for these 50 participants, $p = 0.07$.

²⁰ An exact McNemar's test found no statistically significant difference in the proportion of 'yes' responses to the free will (63%) and moral responsibility questions (65%) for these 54 participants, $p = 1.0$.

6.3 Methods for Study 5

In previous studies, vignettes specified that, '[scientists] will never be able to predict whether you will decide to eat a banana today or when you will eat it if you do.' However, in this study, the vignette instead specified that, 'they will never be able to predict whether you will decide to cheat in a game of poker with friends or when you will cheat if you do'. By making this slight change, we hoped to help participants conceptualize moral violations as harms without raising the emotional valence of the vignette. Other than this change in the vignette, this experiment was the same as study four, including the same moral responsibility probe.²¹

6.4 Results and Discussion of Study Five

As in previous studies, participants who failed either of two comprehension controls appearing at the beginning of this study were excluded from our analyses. The results for the remaining 101 participants were subjected to McNemar's test. As in studies 2 and 3 (but not study 4) participants responded to the free will and moral responsibility questions in significantly different ways, as shown in Figure 7.²²

Order effects also patterned with previous studies. Those participants who received the free will question first were significantly more likely to offer 'yes' responses to the moral responsibility question than to the free will question, while there was no difference in 'yes' responses for the two questions when the moral responsibility question appeared first.²³

7. Our Results in Context

Our results may be compared to other experimental results, although direct comparison can be misleading. For example, if one assumes a connection between

²¹ Participants were located in the United States. One-hundred ten participants took part in this study. Seventy-one percent of participants identified themselves as male and 29% as female. Twenty-four percent of participants were aged 18–24, another 53% were 25 to 34, and 99% of participants were under 65. Fifty-two percent of the participants had not attained a four-year college degree, and 91% had not gone beyond such a degree in their education. (All participants responded to the gender, age, and education questions.) Participants were paid \$0.40 for their participation. Participants were prohibited from participating more than once in this study.

²² An exact McNemar's test determined that there was a statistically significant difference in the proportion of 'yes' responses to the free will (63%) and moral responsibility (73%) questions, $p = 0.006$.

²³ An exact McNemar's test determined that there was a statistically significant difference in the proportion of 'yes' responses to the free will and moral responsibility questions for those 48 participants who responded to the free will question first, $p = 0.039$. Twenty-nine responded 'yes' to the free will question, 36 responded 'yes' to the moral responsibility question. On the other hand, an exact McNemar's test found no statistically significant difference in the proportion of 'yes' responses to the free will and moral responsibility questions for those 53 participants who received the moral responsibility question first, $p = 0.25$. Thirty-five responded 'yes' to the free will question, 38 responded 'yes' to the moral responsibility question.

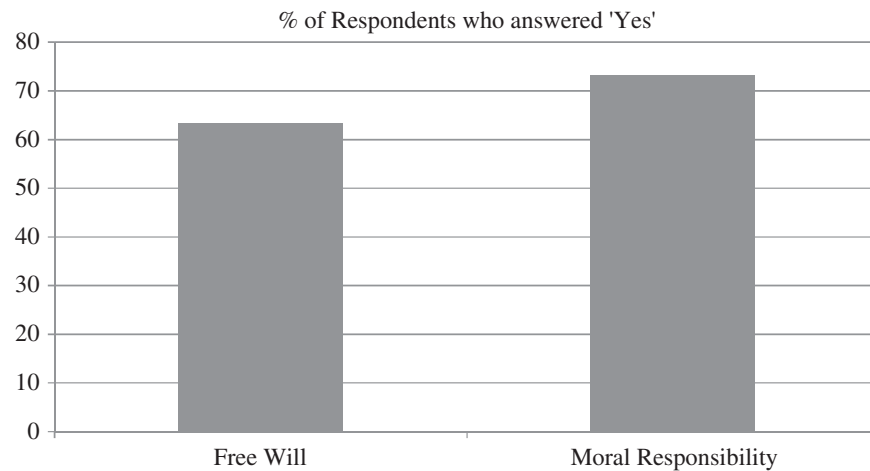


Figure 7 Percent agreement with free will and moral responsibility attributions in the 3rd within-participants, forced-choice neuroscience*concrete*not predictable study (cheating at poker, 'are we morally responsible').

predictability and determinism, our results in study 1 seem to uphold the claim that people are compatibilists about moral responsibility in concrete situations but incompatibilists about moral responsibility in abstract ones, consonant with Sinnott-Armstrong and Nichols and Roskies. The idea that we might hold each other responsible even if we aren't free on a libertarian conception of freedom (e.g. because neuroscience shows that causal determinism is true) is usually interpreted as a compatibilist position—it maintains the necessity of freedom for responsibility by construing freedom as absence of coercion. But the problem with this comparison, and with interpretation of our results in traditional terms, is that it is unclear why such judgments should be classified as compatibilist or incompatibilist if determinism is not directly manipulated. Moreover, *pace* Roskies and Nichols (2008), who see their pattern of responses as 'consistent with what would be expected if judgments of moral responsibility, blameworthiness and freedom were correlated', we found that only the first two dependent variables were correlated across concrete/abstract conditions. This suggests that assessments of freedom (on any conception of freedom) did not drive responsibility judgments, as the necessity claim predicts and as tradition leads us to expect.

An overall assessment suggests instead that the appropriateness of interpreting experimental results from the perspective of traditional philosophical debates should not be taken for granted. To say free will and responsibility judgments are not fully—or even primarily—driven by beliefs about determinism, is not to say that these beliefs play no role in the full explanation of free will and moral responsibility judgments. Indeed, that factors such as concreteness and affect can sway moral responsibility and free will judgments in different ways suggests that they are outcomes of processes that operate independently across a large number of factors. For example, Knobe (2014, p. 5) has suggested that moral responsibility judgments operate more or less independently of causal determinism, but proposes that these

judgments differ in relation to a concrete/abstract factor. The current studies suggest that such judgments may also be sensitive to factors such as domain of explanation.

The fact that in our study moral responsibility and free will judgments diverged without priming by a determinism factor suggests that it may be experimentally fruitful to approach these questions without taking the compatibilist/incompatibilist framework for granted in research design. Once we adopt this perspective we might consider determinism one factor among many that may affect moral responsibility and free will judgments without presupposing either that determinism is a major influence on folk judgments of either kind or that the two kinds of judgments will necessarily co-vary in relation to any of the factors. This would be facilitated by an alternative general framework in which to develop such studies. We see Mandelbaum and Ripley's (2012) NBAR account of patterns of judgments of moral responsibility as a promising start to developing this alternative.

8. Connection to Mandelbaum and Ripley's NBAR Proposal

The NBAR hypothesis holds that we have an entrenched belief (or inference mechanism) that if a Norm is Broken, an Agent is Responsible, and that moral responsibility judgments are tied to norm-breaking. Mandelbaum and Ripley (M&R) propose that this mechanism is more engaged by concrete as opposed to abstract scenarios because concreteness makes norm violation more salient. Concreteness gets linked to more responsibility (and abstract to less) because the NBAR mechanism is more activated in concrete scenarios. Moreover, when subjects have been primed with deterministic scenarios to judge that agents are not responsible, the entrenchment of NBAR explains why the cognitive dissonance in concrete scenarios—are they responsible, or not?—is resolved in favor of holding them responsible.

We find it promising that the NBAR hypothesis makes no essential reference to free will, determinism, or any specific psychological states (including affect). The hypothesis posits a mechanism whereby expectation-violations are linked to responsibility judgments. This draws a direct connection between one factor (norm-breaking) and one kind of judgment (responsibility) without entailing anything about the relation between responsibility and free will or responsibility and a determinism factor. However, we suggest two modifications to the hypothesis that would demonstrate more clearly the new research designs it can promote. One involves generalizing the consequent. The other involves eliminating the entrenchment feature, which is an artifact of the traditional compatibilist/incompatibilist framework.

First, what M&R mean by 'norm' is extremely weak, and is (as they make clear) closer to 'expectation' in a very broad sense: a norm violation is 'a deviation from the normal course of events', such that 'each of the beliefs one has about how the world ought to be (in the broadest possible sense of "ought") counts as a norm.' A norm is just a regularity or an expectation of uniformity.

The generalization is motivated by the fact that regularities, as defined above, are ubiquitous, while agents are not. That is, it is unlikely that the concept of agency

provides a level of generality for NBAR's consequent that is appropriate to the concept of a norm in the antecedent. One way to provide this generality would be to replace the concept of agency with a notion of a cause, whether animate or not; human agents would be special cases. For example, Mackie's (1965) concept of an INUS condition—an Insufficient but Necessary part of a Unnecessary but Sufficient condition—captures the idea that a cause is something such that if it had not occurred or been present in the circumstances the effect would not have occurred. Thus, to borrow an example from M&R, a storm-blown tree is responsible for killing one's first-born because it is an INUS condition for the baby's death even if it is not an agent. (We consider Mackie's proposal promising, but are not committed to this precise way of generalizing NBAR.)

The generalized form of NBAR, on this suggestion, is 'Regularity Broken, Find Cause' (RBFC), which will have moral norms and agents as special cases of regularities and causes (e.g. INUS conditions) respectively. RBFC claims that we are expectation or regularity conservatives—we are biased to keep our expectations (or assume our regularities are correct) rather than to cease having them or change them. When something surprising happens, our first impulse is to identify a cause of the disruption rather than assume we did not grasp the real regularity or that our expectations were not justified. We treat the surprise as a violation, not an instance of another regularity.

From this perspective, our study 1 was in effect a test of RBFC. By manipulating predictability, we manipulated the ability of scientists to identify INUS conditions by describing situations where norms are very reliable (high accuracy of prediction) or not. Subjects' responsibility judgments tracked those conditions in which INUS conditions could be easily identified even when their free will judgments did not. In this way, RBFC provides a framework for formulating specific hypotheses regarding responsibility and, along with other hypotheses involving other factors and the relation of these factors to free will judgments, can contribute to experimentally verifying or falsifying the assumed necessary connection.

Determinism, on the other hand, is relevant to explaining why a cause came about. This distinction can explain why subjects' responsibility judgments vary in some conditions (e.g. concrete/abstract, high/low affect) even when subjects are primed with deterministic or indeterministic scenarios. Instead of saying (as M&R do) that people are intuitive incompatibilists but that the incompatibilism gets jettisoned because NBAR conflicts with it and is more entrenched, we suggest a simpler link. On our view, processes for identifying causes ('Who/what did it?') and explaining them ('Why did [who/what] do it?') operate largely independently and can be manipulated independently. Participants asked to assess responsibility are being asked to identify causes, while free will judgments are responses to requests for an explanation. Thus, whether primed with determinism or not, subjects presented with a broken norm or violated expectation are biased to identify a cause (as RBFC hypothesizes). This need not be accompanied by an explanation. There is no need to invoke entrenchment because a cause judgment doesn't have to triumph over

an explanation judgment. The problem that entrenchment is posited to solve is an artifact of the traditional framework.

Similarly, Nichols and Roskies suggest that the intuition that we are morally responsible is non-negotiable: if determinism is true, judgments are compatibilist, and if indeterminism is true, then judgments are incompatibilist. On our view, if responsibility judgments hold firm whatever the case may be metaphysically (with regard to determinism), this is just another way of saying they are independent of the explanations we might give for these causes.

This suggests that, at least in some cases, rather than free will being necessary for moral responsibility, the necessity relation may run in the other direction. In terms the folk might use, 'He did it of his own free will' entails 'He did it'. Ascribing free will to an agent as an explanation of what he did can be, for some subjects, sufficient for having identified that agent as the cause—the responsible party. It follows that being identified as the cause is necessary for ascribing free will to an agent as an explanation for his having caused what he did. He may not have done it freely. But he must have done it to be free.

9. Summary

To date, experimental philosophy in this area of moral psychology has been driven by the compatibility question: are free will and moral responsibility compatible with determinism? From this perspective, when free will and moral responsibility judgments do not move in tandem, either the compatibilism or incompatibilism (or the switch from one to the other) must be explained, often by positing an additional factor that interacts with the determinism factor. Our results suggest that research designs that draw on traditional philosophical assumptions about this relation may build in certain conceptual or inferential links that ordinary folk judgments do not share and yield results that are interpreted in ways that may not be accurate.

On our view, the basic research questions are: what factors explain judgments of moral responsibility, and what factors explain judgments of free will? And given answers to these questions, what will this reveal about the relation between these two kinds of judgments? Sommers has also suggested (2010, p. 209) moving away from the compatibilist question to focusing on factors affecting free will and moral responsibility judgments directly. We see our alternative as a step in this direction.

Carrie Figdor
Department of Philosophy,
University of Iowa

Mark Phelan
Department of Philosophy,
Lawrence University

Appendix

Study 1 contained versions of the following questions (appropriate to each scenario) to measure subjects' judgments.

Please indicate your level of agreement with the following statements (scale of 1–6, 6 = strongly agree, 1 = strongly disagree). (An 'I don't know' option will also be available.)

[These are samples of each question. Each will be altered to fit each scenario.]

1. If the neuroscientists are right, our decisions are up to us.
2. If the neuroscientists are right, we are able to make decisions of our own free will.
3. If the neuroscientists are right, we should be held morally responsible for our actions.
4. If the neuroscientists are right, we deserve to be blamed for our bad actions.
5. If the neuroscientists are right, we deserve to be praised for our good actions.

References

- Feltz, A. and Cokely, E. 2009: Do judgments of free will and moral responsibility depend on who you are? Personality differences in intuitions of compatibilism and incompatibilism. *Consciousness and Cognition*, 18, 342–50.
- Frankfurt, H. 1969: Alternate possibilities and moral responsibility. *Journal of Philosophy*, 46, 829–39.
- Knobe, J. 2014: Free will and the scientific vision. In E. Machery and E. O'Neill (eds), *Current Controversies in Experimental Philosophy*. Abingdon: Routledge.
- Mackie, J. 1965: Causes and conditions. *American Philosophical Quarterly*, 2(4), 245–64.
- Mandelbaum, E. and Ripley, D. 2012: Explaining the abstract/concrete paradoxes in moral psychology: the NBAR hypothesis. *Review of Philosophy and Psychology*, 3, 351–68.
- McKenna, M. 2009: Compatibilism. In E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Winter 2009 Edition, URL = <<http://plato.stanford.edu/archives/win2009/entries/compatibilism/>>
- Nahmias, E., Morris, S., Nadelhoffer, T. and Turner, J. 2005: Surveying freedom: folk intuitions about free will and moral responsibility. *Philosophical Psychology*, 18(5), 561–84.
- Nahmias, E., Morris, S., Nadelhoffer, T. and Turner, J. 2006: Is incompatibilism intuitive? *Philosophy and Phenomenological Research*, 73, 28–53.
- Nahmias, E., Coates, D.J. and Kvaran, T. 2007: Free will, moral responsibility, and mechanism: experiments on folk intuitions. *Midwest Studies in Philosophy*, 31(1), 214–42.
- Nichols, S. and Knobe, J. 2007: Moral responsibility and determinism: the cognitive science of folk intuitions. *Noûs*, 41(4), 663–85.

- Roskies, A. and Nichols, S. 2008: Bringing moral responsibility down to earth. *Journal of Philosophy*, 105 (7), 371–88.
- Sinnott-Armstrong, W. 2008: Abstract + concrete = paradox. In J. Knobe and S. Nichols (eds), *Experimental Philosophy*. New York: Oxford University Press.
- Sommers, T. 2010: Experimental philosophy and free will. *Philosophy Compass*, 5(2), 199–212.
- Strawson, P.F. 1974: *Freedom and Resentment and Other Essays*. London: Methuen.
- Van Inwagen, P. 1983: *An Essay on Free Will*. Oxford: Clarendon Press.
- Vihvelin, K. 2011: Arguments for incompatibilism. In E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), URL = <<http://plato.stanford.edu/archives/spr2011/entries/incompatibilism-arguments/>>
- Zagzebski, L. 1991: *The Dilemma of Freedom and Foreknowledge*. New York: Oxford University Press.