On the Very Concept of Free Will

Joshua May

In *Synthese* vol. 191, no. 12 (2014), pp. 2849-2866

[Penultimate draft; citations should be to the final version at springer.com]

Abstract: Determinism seems to rule out a robust sense of options but also prevent our choices from being a matter of luck. In this way, free will seems to require both the truth and falsity of determinism. If the concept of free will is coherent, something must have gone wrong. I offer a diagnosis on which this puzzle is due at least in part to a tension already present in the very idea of free will. I provide various lines of support for this hypothesis, including some experimental data gathered by probing the judgments of non-specialists.

Total Word Count (w/references, notes, etc.): 9,731

Keywords: freedom, moral responsibility, experimental philosophy, incompatibilism, compatibilism, cluster concept, prototype

1. Introduction

Debates about free will seem to pull rational people in multiple directions. One can readily see a deep tension in our thinking on this topic when considering the problem in terms of determinism—the thesis that, given the past and the laws of nature, there is only one possible future. There is a well-known mystery here: the existence of free will seems to require both the truth and falsity of determinism. Two famous arguments can be used to motivate such a conclusion.

On the one hand, if determinism is true, then we lack a robust sense of options, given the past and laws of nature. As Peter van Inwagen (2000) has made vivid with his "consequence argument," it looks like we can't do otherwise than what we do, assuming we can't change the past or the laws of nature. Yet this seems required if we are to possess free will. So the falsehood of determinism appears to be a necessary condition for free will.

On the other hand, the falsity of determinism gives rise to what can be called the "luck argument." Indeterminism simply makes more than one future compatible with the past and the laws, yet this seems only to introduce an element of luck into whether you, say, choose the soup or the salad with dinner. After all, if which action will occur is not completely determined by your character, psychological states, circumstances, and so on, then the outcome seems at least partly a matter of chance (cf. Mele 2006). So the truth of determinism appears to be a necessary condition for free will.

One of these arguments may ultimately be unsound, but I do not aim to add to the debate by adjudicating that issue. Rather, the point to register here is simply that these arguments each have an intuitive force and thus illustrate our puzzle: it may at least seem

that the very idea of free will is simply incoherent or, at any rate, impossible to satisfy (cf. Nichols 2006). P. F. Strawson (1962) called this the position of "the genuine moral sceptic." But, as Strawson recognized, belief in free will and moral responsibility is difficult to shake. There seem to be both merely psychological and rational reasons for this, but we needn't rehearse them here. Let us just say, if we are to avoid skepticism about free will, how are we to deal with this puzzle?

2. Two Concepts: Liberty & Ensurance

A familiar way of proceeding treats the problem of free will as a mere "verbal dispute." That is, there are two properties each dubbed "free will" by different authors. One is incompatible with determinism, while one of them requires it. For purposes of clarity, we should simply eliminate the term "free will" from our lexicon, and replace it with others, such as "free will$_1$" and "free will$_2$" (cf. the discussion in Chalmers 2011). Once this is done, there is no point asking whether "free will" is possible in a deterministic or indeterministic world. This question rested on a false presupposition, namely that "free will" is not ambiguous. And this may help to explain why debates about free will seem to result in what some have appropriately called "dialectical stalemates" (e.g. Fischer 1994, pp. 83–5), in which neither side seems able to defend their position without begging the question against their opponent. This treatment of free will dates at least to Hume, but it has some recent adherents as well (e.g. Balaguer 2010, esp. ch. 2).

I don't find this position plausible. For one thing, it seems uncharitable: though the problem of free will may in some sense ultimately rest on a confusion, it seems unlikely that the confusion is as simple as this (cf. van Inwagen 2008). Moreover, it is empirically unmotivated: when we have a single term in our lexicon, our default should be to treat that term as not ambiguous. If an ambiguity hypothesis is introduced simply because it will make a philosophical problem go away, then the hypothesis hasn't been given sufficient support to defeat the presumption of non-ambiguity.

The alternative proposal I defend also locates a certain measure of confusion in the traditional dispute, but it occurs at the psychological rather than the semantic level. That is, "free will" (and its cognates, such as "freely") may well be unambiguous in natural language, but I suggest that their application is guided by at least two distinct factors. When both appear to be present, free will is judged to be present. When both factors appear to be absent, free will is judged to be absent. When one but not the other factor is present, it is unclear whether or not free will is present.[1]

This duality in our concept of free will goes mostly unnoticed, and causes little trouble, because the vast range of cases that we are familiar with in everyday life seem to be ones in which both or neither of these factors are present. It is only certain esoteric cases and certain philosophical arguments that force us to confront the question of what happens when exactly one of these factors is present. And it may be that the application

[1] This proposal is inspired by discussions with John Maier and by the treatment of weakness of will in May & Holton (2012).

conditions for "free will" do not fix any determinate answer to this question. Thus our perplexity when confronted with such cases, and with such arguments.[2]

What, then, are these two factors? The first let's dub "liberty." An agent has *liberty* in a situation just when she has at least two genuine options for action in that situation. Liberty is intuitively important for acting freely since lacking options seems to render choice an illusion. When theorists take this factor to be relevant for free will, the falsehood of determinism appears to be a necessary condition. Let's call the second factor "ensurance." An agent has *ensurance* with respect to an action just when the action depends in an appropriate way on her mental states and her environment. Ensurance captures the kind of control that seems important for agents who act freely and responsibly. When theorists focus on this factor in their account of free will, the truth of determinism appears to be a necessary condition.[3]

Now, these factors are purposefully defined here in a general way, leaving them open to a more specific characterization. For example, I have not precisely defined "option," much less "genuine option." The same goes for phrases such as "depends in an appropriate way." Compatibilists and incompatibilists have notoriously fought over specific ways of characterizing such notions, which then conflicts with their opponent's view. But that is not our goal here. I want first and foremost to provide an account of ordinary thinking about free will, which will naturally lack the specificity of a philosophical analysis.

These two factors, liberty and ensurance, each play a role in application of the concept of free will. However, I propose they are not individually necessary and jointly sufficient conditions, as a classical, definitional account would maintain. However, I do propose that, at least as far as ordinary thinking goes, liberty and ensurance are together (or perhaps with other factors) sufficient for appropriate application of the concept of free will. This might seem impossible if liberty and ensurance are, upon theoretical reflection and refinement, best understood as incompatible. After all, something like the luck argument seems to show that a certain kind of control (which one might assume is identical with ensurance) is impossible in an indeterministic universe (and one might assume liberty is only secured in such universes). But we have not made the parenthetical assumptions regarding the ordinary concept of free will. We have not said that ensurance is precisely the kind of control ruled out in an indeterministic universe or that one has liberty only if determinism is false. It is only if these general notions are so specified by theorists in the free will debate that we are lead to the mystery surrounding free will.

So application of the concept of free will is relatively straightforward when both factors are apparently either present or absent, but not so when only one factor is missing.

---

[2] Throughout I will speak of "our" concept of free will and the intuitions "we" have, but this is primarily for convenience and should of course be taken with a grain of salt. It is unclear how far we can generalize from studies using participants largely from Western cultures (although, for a relevant cross-cultural study, see Sarkissian et al 2010).

[3] For a discussion of options and their relation to freedom, see John Maier (forthcoming). The labels "liberty" and "ensurance" are his invention, although he doesn't employ them in that paper. At any rate, these should be thought of as technical terms, even though they may be used elsewhere in the free will literature with different meanings. In particular, "liberty" has a long history and some have used it to simply designate what we now call "free will" (e.g. Hume and Reid).

To make sense of this, we must at least treat the concept of free will as non-classical. There are several alternative models, and I would like to remain as ecumenical as possible, embracing those that fit with the intended explanation of the mystery with which we began.

One model that has gained some traction in the theory of concepts is a prototype theory, according to which each factor (e.g. liberty and ensurance) plays a contributory role. If enough of the prototypical features are met, then that will be sufficient for the item in question to fall under the concept. So we can make sense of some jointly sufficient conditions for application of the concept, but this will not necessarily entail that subtracting one of the factors in the bundle will leave a cluster that makes for a sufficient condition. Each factor's contribution toward appropriate application of the concept will be probabilistic. On most prototype theories this probabilistic relation will be determined entirely by how many features the item in question (e.g. chess) shares with prototypical instances of something falling under the extension of the concept (e.g. football for the concept *game*).

Experimental support for non-classical theories, like the prototype theory, involves "typicality effects." For example, people can easily rank items, such as fruits, as being more typical than others—e.g. an apple is a more typical fruit than a fig. Similarly, we can more quickly recall and categorize items the more typical they are.[4] These effects can support either a *prototype theory*, which holds that the relevant concepts are applied using a set of features of a prototype. But another contender is *exemplar theory*, which involves resemblance to particular paradigm examples of the given type. For simplicity, I'll focus on the prototype theory as a model, but any non-classical theory might work for our purposes.

Most of the empirical research on typicality effects has been on concrete objects, rather than abstract ideas, such as freedom. But there is some evidence that abstract concepts, such as *science* and *crime*, exemplify typicality effects, and there may even be evidence for some moral concepts (see Park 2013 for a recent discussion). So applying a non-classical model to free will is not outlandish, and we will see that it nicely captures the data to be reported below. Moreover, while I have suggested that in paradigm cases of acting freely we have both liberty and ensurance, this needn't rely on a theory focused solely on *similarity* with a prototype or exemplar. I minimally claim that these two factors influence application of the concept without necessarily contributing to an analysis in terms of necessary and sufficient conditions. Hence, I'll attempt to remain more neutral and distinguish this proposal from others in the literature on concepts by calling it a "cluster theory." Of course, any theory of concepts seems subject to formidable objections; a cluster theory is merely one of various models that could apply to free will.

On this theory, the ordinary concept of free will is not incoherent but may seem so when seeking a classical analysis. For the ordinary person on the street, whether someone acts freely is determined in part by whether, roughly put, that person has options and is in control of the relevant action. Theorists might argue for an incompatibility here, but this relies on two further claims. First, we'd have to say that liberty is a necessary condition

[4] Much of this work is done by Eleanor Rosch (e.g. Rosch 1975). See Laurence and Margolis (1999) for a philosophical overview of this experimental work as well how various theories of concepts relate to it.

on free will and is incompatible with determinism. Second, we'd have to say that ensurance is a necessary condition for free will and is incompatible with indeterminism. On the cluster theory, as far as ordinary thinking goes, neither of these factors is a necessary condition for free will and no explicit connection is made with the theoretical concept of determinism. Yet, like the verbal dispute theory, this may help to explain why the theoretical debates seem to easily become locked in dialectical stalemates, as both sides are highlighting an important factor in the ordinary concept.

As I have emphasized, though, this proposal is importantly different from the verbal dispute theory. First, it is not as uncharitable: I do not think that philosophers have been "talking past" one another, but rather that they have been talking with one another using a terminology whose application conditions are complex and subtle. Moreover, this theory is quite conciliatory, suggesting that both sides of the debate—compatibilists and incompatibilists—are partially right. The application conditions for this concept are non-classical and both sides have emphasized an important factor. So we needn't attribute a simple confusion or error to either theorist. Second, and more fundamentally, it is not a thesis about semantics. It is thus compatible, unlike the verbal dispute view, with the idea that "free will" and its cognates are not ambiguous. It is also compatible with the idea that "free will" and its cognates are semantically invariant, in the sense of not having their application conditions affected by the context of utterance (cf. Hawthorne 2001).

Moreover, the cluster theory makes some fruitful empirical predictions. In particular, it predicts that when both factors appear to be present, we will tend to ascribe free will, when both are absent, we will not be so inclined, and that when exactly one of these factors is present, some measure of confusion will reign. And, as the studies described in what follows show, this is precisely the pattern of ascriptions that we do in fact find.

3. Some Previous Empirical Work

The cluster hypothesis is at least partly an empirical one, as it makes a claim about our ordinary thinking. Thankfully, there is a considerable amount of work being done on ordinary intuitions about freedom and responsibility. In fact, the literature is exploding, but the story so far is roughly this. Initial results suggested that non-specialists are inclined to make compatibilist judgments, in that most will count someone as responsible and acting freely even in a universe that is described (in ordinary terms) as deterministic (e.g. Nahmias et al. 2006). However, sometimes these compatibilist majorities are not overwhelming. This is especially so for judgments about free will in the more neutrally described deterministic universe—i.e. something like a "rollback" scenario in which a universe plays out exactly the same way over and over, starting with the same laws of nature and initial conditions.

Moreover, later data indicate a more complex picture, according to which people will tend to have incompatibilist intuitions regarding an abstract case but not a concrete one (Nichols & Knobe 2007). In one experiment, subjects were presented with, among other things, a deterministic universe:

> Imagine a universe (Universe A) in which everything that happens is completely caused by whatever happened before it. This is true from the very beginning of

> the universe, so what happened in the beginning of the universe caused what happened next, and so on right up until the present. For example one day John decided to have French Fries at lunch. Like everything else, this decision was completely caused by what happened before it. So, if everything in this universe was exactly the same up until John made his decision, then it *had to happen* that John would decide to have French Fries. (p. 669)

Participants in the concrete group then read about the case of Bill who lives in this universe and kills his family in order to be with his secretary. In line with earlier results, they overwhelmingly said Bill was morally responsible, yielding compatibilist intuitions. However, participants in the abstract group didn't read about Bill and were simply asked: "In Universe A, is it possible for a person to be fully morally responsible for their actions?" The vast majority gave the *incompatibilist* response ("No").

Nichols and Knobe conducted a further experiment the results of which suggest that it is precisely the emotionless nature of a case that yields the incompatibilist judgments. Using only concrete cases, they constructed four vignettes that varied two factors systematically: determinism (the agent is in deterministic Universe A vs. indeterministic Universe B) and affect (the agent rapes a stranger vs. cheats on his taxes). Statistical analyses indicated that the emotional aspect of a case significantly affected participants' responses regarding the deterministic universe, such that those in the high- rather than low-affect group were more inclined to say the agent was responsible.

On the face of it, these data may seem to support the cluster hypothesis, given that our ordinary thinking can reflect both compatibilist and incompatibilist reactions. I have suggested that such reactions are tied to liberty and ensurance, so the cluster hypothesis is compatible with any theory that refrains from discounting either compatibilist or incompatibilist intuitions as due to a "performance error" or otherwise not reflecting our ordinary concepts. But Nichols and Knobe tentatively suggest that we should consider the incompatibilist responses as reflecting our true concept of responsibility since they aren't distorted by emotion. According to their affective error theory, as we might call it, compatibilist judgments are simply due to a performance error, distorted so that they do not reflect one's conceptual competence.

However, this conclusion has come under fire for reasons we needn't belabor here (for a summary, and recantation, see Knobe 2014, n. 2). What remains is converging evidence from a number of researchers showing that our ordinary thinking is both compatibilist and incompatibilist, but what processes drive these intuitions is still uncertain. Perhaps we should have expected this state of affairs, however, for there is also some existing empirical support for a hypothesis which posits factors like ensurance and liberty that both affect intuitions about free will (and responsibility). Before turning to my own studies, let's briefly discuss such antecedent sources of empirical support.

First, in an early study, Woolfolk, Doris, and Darley (2006/2008) tested whether judgments about moral responsibility were affected by one's situational constraints as well as one's level of identification with the action. Their key case involved a man, Bill, who learns that his wife has been unfaithful with another man, Frank. Bill soon has the opportunity to kill this man, but based on the unrelated demands of hijackers who have come to occupy their airliner and want to display their ruthlessness. In all versions of the case, Bill kills Frank, ostensibly at the demand of the hijackers. In different versions of

the scenario, however, Woolfolk and colleagues varied how reluctant or resolved Bill was to kill Frank (identification); they also varied the likelihood of successfully resisting the demand of the hijackers if Bill tried to do so (situational constraint). The results across several experiments were that both identification and situational constraint affected participants' judgments about the degree to which Bill was morally responsible for killing Frank.

The authors' primary goal was to show that ordinary people will tend to hold someone responsible even when the agent couldn't have done otherwise, which is friendly to compatibilism. However, their data provide some initial support for the cluster hypothesis in that both the agent's options and psychological control affected people's judgments. In fact, Woolfolk and colleagues conclude that ordinary thinking about responsibility is "contextualist" by which they mean in part that "differing considerations are salient to moral responsibility attribution in different contexts" (p. 77; cf. also "variantism" in Knobe and Doris 2010). This theory is not fully fleshed out, but it no doubt has some affinity with mine. A key difference of course is that they do not specifically posit ensurance and liberty as key factors and their focus is moral responsibility. But the cluster theory is open to the possibility that other factors may be playing a role as well and that the factors at play will affect thinking about both moral responsibility and free will.

There is some relevant evidence for the latter claim in the final experiment from Woolfolk et al. While the primary goal of the previous two studies was to measure attitudes about responsibility, the authors wanted to make sure they manipulated situational constraint enough so that participants believed the agent couldn't have done otherwise. In this third study, the researchers asked a number of freedom-relevant questions about the same vignettes but without manipulating identification. The questions included whether Bill "was constrained," "was forced," "was free to do other than he did," or "had a choice." The results were roughly the same, suggesting that judgments of free will can likewise be affected by the degree to which an agent's options are limited in addition to the degree to which the agent identified with the action. While the constraints on one's choices are not as extreme as they are in a deterministic universe, we have evidence that liberty influences ordinary judgments about freedom and responsibility.

A final preliminary line of support for my hypothesis can be found in a more recent study. Feltz, Perez, and Harris (2012) attempted to determine whether those providing compatibilist responses would be more likely to provide explanations in terms of the agent's psychological states ("decision explanations"). They predicted that, on the other hand, those providing incompatibilist judgments would be more likely to provide explanations of the agent's action in terms of non-psychological phenomena, such as "events completely out of his control" ("causal explanations"). Despite the perhaps misleading label, such explanations plausibly include appeal to the constraints on one's options that limit one's liberty. In five studies, Feltz and colleagues tracked intuitions about an agent's freedom and responsibility regarding an action and then asked why the agent did it. Explanations were categorized as either "decision" or "causal" using two independent coders. The results across all five studies support their key hypothesis that explanations employing psychological states tended to appear in those registering compatibilist responses, while incompatibilist responses tended to go with explanations that referred to factors outside of the agent's psychological control. Again, such data fit

well with my proposal that factors like ensurance and liberty are tied to compatibilist and incompatibilist thinking among ordinary people.[5]

4. Experimental Support

While the research on ordinary judgments about freedom and responsibility has been complicated and controversial, it seems we are roughly back where we started with Nichols and Knobe's original study. That is, ordinary thinking, like philosophical thinking on this issue, is mixed. Moreover, there is some reason to think that these mixed results are at least partly explicable in terms of liberty and ensurance.[6] But no previous study provides clear and direct evidence for the cluster hypothesis, so I decided to go forth and seek further and more direct empirical support. If true, it could explain, not only the dialectical stalemate one sees in the philosophical literature, but the mixed results among ordinary folks as well.

4.1 Experiment 1: Brainwashing

To this end, I designed four simple cases that varied ensurance and liberty, yielding a factorial design:

Table 1: Experimental Design

	Ensurance	**No Ensurance**
Liberty	*Cell 1*	*Cell 2*
No Liberty	*Cell 3*	*Cell 4*

The cluster hypothesis predicts that ordinary non-specialists would be highly inclined to say a person acted freely if she had both liberty and ensurance. Conversely, people should be highly inclined to say an agent didn't act freely if she lacked both factors. In Cells 2 and 3, however, the prediction is that the results would be mixed, without a clear consensus among participants about the agent's freedom with respect to her action. Moreover, the cluster hypothesis predicts that a statistical analysis would show that both factors played a causal and substantial role in ordinary judgments about free will. Contrary theories, which place emphasis on only liberty or only ensurance in ordinary

[5] There is further evidence for the general idea that multiple factors drive ordinary judgments about freedom and responsibility (e.g. Feltz et al. 2009; Knobe and Doris 2010; Weigel 2011). But these factors may not nicely align with ensurance or liberty. The cluster hypothesis is not necessarily in conflict with these proposals, however.

[6] Of course, perhaps a driving force behind the two intuitions is a difference between abstract and concrete scenarios (Sinnott-Armstrong 2008). There is some evidence for this, as it appears to be what remains from Nichols and Knobe, and the basic result has been replicated, including cross-culturally (Sarkissian et al. 2010). This account is in principle compatible with the cluster hypothesis, however. In fact, it may well be that more abstract cases make liberty more salient while concrete cases do so for ensurance. Alternatively, abstractnesss could be a further factor. I remain neutral on the merits of the abstract-concrete theory, although further research could determine whether it is incompatible with a cluster account.

thinking about free will, should expect that their preferred factor would be much more predictive of participants' attributions of free will.

Using this design I conducted an experiment that systematically manipulated liberty and ensurance. I attempted to operationalize liberty in a way that's consistent with previous research, which focuses on a robust notion of options that are plausibly inconsistent with determinism. I opted for using a modification of the various rollback universes. Ensurance was more difficult to undermine in such drastic terms, yet it should be similar in degree to the lack of liberty if we are to compare their impact on attributions of freedom.[7] Previous research has shown that one way to reduce ordinary attributions of free will is to impair the agent's psychological capacities through something like brainwashing (e.g. Mele 2014, sect. 1). So I attempted to more strongly undermine ensurance in the relevant vignettes through the ordinary notion of "brainwashing."

Such manipulation may seem an entirely separate issue from one's actions depending in an appropriate way on one's mental states and environment (ensurance). But there is evidence that whether we tend to think an agent acts freely when manipulated depends precisely on whether this impairs psychological capacities, either through (a) developing a discordance among one's mental states or (b) corrupting one's access to information required to make a decision (Sripada 2012). An increase in one or both of these factors due to manipulation should undermine ensurance, understood again as having one's action depend in an appropriate way on one's mental states and environment.[8]

This led to designing four vignettes that fit the factorial design, which I randomly assigned to ordinary non-specialists on the topic of free will.[9] Each vignette consisted of two short paragraphs, which subjects were instructed to read carefully as they would be answering questions about it shortly after. For each cell, the first paragraph was designed to either setup a case involving liberty or lacking it (of course, no words were in bold for participants):

---

[7] In a pilot study, I attempted to manipulate ensurance by describing the protagonist as having an "uncontrollable urge" but this didn't fully work. Liberty had a statistically significant effect, but not ensurance, although it did influence responses in the predicted direction. Hence, in the experiment reported here, I attempted to more drastically undermine the protagonist's ensurance along the same dramatic lines in which she lacks liberty. I've opted not to provide the details of this pilot only in the interest of space.

[8] One might worry that when an agent is manipulated this undermines freedom for the same reasons that a deterministic universe does (as proponents of famous manipulation arguments allege). In that case, I wouldn't have operationalized ensurance in a way that is distinct from the operationalization of liberty (tied as it is to determinism being true). But there is some evidence that, for ordinary folks, free will seems undermined by manipulation in a way that is different from determinism (see Feltz 2013).

[9] For all studies in this paper, participants were users of Amazon's Mechanical Turk website in the U.S. and were paid for participating. Mturk is a popular and reliable place for soliciting human subjects for scientific research, and it provides an even more diverse sample than simply using university students (see Buhrmester et al 2011). In each of the studies, the mean age of participants was 30-32 and between 50-60% were male. I took standard measures to encourage and monitor serious participation. The entire project was approved by Monash University's Human Research Ethics Committee (project number CF12/1686 – 2012000918).

> Imagine there is a universe that is re-created over and over again, starting from the exact same initial conditions and with all the same laws of nature. In this universe the same initial conditions and the same laws of nature [**needn't/must**] cause the exact same events for the entire history of the universe. So, every time the universe is re-created, [**some things may not/everything must**] happen the exact same way.

All participants continued reading a second paragraph, which manipulated ensurance and fit with their version of the previous paragraph as involving liberty or not. In Cells 1 and 3, where ensurance was held fixed, participants read one of the versions of the second paragraph (varying liberty), both of which were intended to secure ensurance and reiterate the protagonist's psychological state:

> For instance, in this version (Universe 49) a person named Jill **deliberates and decides** to steal a necklace at a particular time and then steals it, even though she knows it's wrong. But [**sometimes when/every time**] the universe is re-created she **deliberates [yet/and] decides [not] to steal** the necklace at that time.

In Cells 2 and 4, lack of ensurance was held fixed while liberty was again varied depending on which version of the first paragraph participants read:

> For instance, in this version (Universe 49) a person named Jill **is brainwashed to have a powerful urge** to steal a necklace at a particular time and then steals it, even though she knows it's wrong. But [**sometimes when/every time**] the universe is re-created she **is brainwashed again, has the powerful urge, and [yet doesn't] steal[s]** the necklace at that time.

After the presentation of one of the vignettes, participants were presented with the following statement and asked for their response on a 7-point Likert-type scale (non-numerically presented):

> Please tell us whether you agree or disagree with this statement:
>
> "In Universe 49, Jill stole the necklace freely."[10]
>
> O O O O O O O
> Disagree completely In between Agree completely

I opted for focusing only on judgments of freedom rather than responsibility as well to simplify the experiment and to avoid taking a stance on whether the two should be treated differently. However, existing data do seem to suggest that the two types of judgment are rather similar, at least in their application among non-specialists.

I also presented some comprehension checks and filler questions. After reading the vignette, subjects were prompted first to briefly describe in a sentence or two the story they just read and next to indicate their degree of agreement or disagreement with

[10] I also tested whether responses would be different using a slightly different measure: "Jill stole the necklace of her own free will" (we might think of this as Experiment 1b; N = 105). The results were basically the same as before.

this statement: (2) "Jill stole the necklace." These were both meant as comprehension questions. Subjects who did not summarize the story at all or who provided a response lower than 5 for the second question (thus failing to indicate some agreement), were excluded from analysis. There were only 17 such people of 243. In addition to the dependent measure, two filler questions concerned their agreement with: (3) "In Universe 49, Jill had a choice about whether to steal the necklace." (4) "In Universe 49, Jill was in control of stealing the necklace."

Using a scale to measure responses to the freedom question provides a more fine-grained way to view them. We can, for example, see if a factor influences people's reactions to a certain degree even if this doesn't push them from one categorical judgment to another. The scale also offers a midpoint, which avoids a forced choice. Yet we can still group responses into either some level of agreement, disagreement, or "in between" in order to view the data categorically.

The results supported the cluster hypothesis. Regarding the statement about Jill acting freely, Table 2 summarizes the mean levels of agreement in each cell from the 226 participants. There is another way to examine the data as well. Table 3 summarizes the percentage of participants who indicated some level of agreement with the statement that Jill acted freely (i.e. the percentage who provided a response of 5 or higher).

Table 2: Mean Freedom Responses for Experiment 1

	Ensurance	**No Ensurance**
Liberty	6.60 (SD=0.89)	3.98 (SD=2.25)
No Liberty	4.67 (SD=2.32)	3.17 (SD=2.25)

7-point scale with higher numbers corresponding to higher levels of agreement with the claim "In Universe 49, Jill stole the necklace freely." N = 226.

Table 3: Proportion in Agreement with the Statement

	Ensurance	**No Ensurance**
Liberty	96%	39%
No Liberty	49%	30%

Simply looking at these descriptive data, we can see that there is a strong tendency for participants to agree that Jill acted freely when both liberty and ensurance appear to be present, while the opposite is true when both factors are absent. But, before drawing any conclusions, we should of course probe further by analyzing the data to see which of either of these factors made a statistically significant difference to responses across these four cells. It turns out that both factors influenced responses. And the effect sizes are quite large, so the differences these factors made were substantial.[11]

Two worries might arise at this point. First, one might insist that ensurance was poorly operationalized; it is really manipulation that is affecting attributions of free will.

[11] I subjected the data to a 2 (E vs. ~E) by 2 (L vs. ~L) between-subjects analysis of variance (ANOVA). There was a main effect of Ensurance, $F(1, 222) = 58.03$, $p < .001$, partial eta-squared = .207, and a main effect of Liberty, $F(1, 222) = 25.6$, $p < .001$, partial eta-squared = .103. There was a significant interaction effect, $F(1, 222) = 4.3$, $p = .040$, partial eta-squared = .019, such that the differences between responses when Liberty was present or absent was greatest when Ensurance was present.

Earlier I pointed to previous research that suggests attributions of free will are undermined by manipulation *via* lack of ensurance (Sripada 2012; cf. Feltz 2013). But we can now provide even further support for this by examining the two "filler questions." These can also serve as checks on the operationalization of ensurance since they concerned the agent's control and choice, both of which plausibly affect whether an action depends in an appropriate way on an agent's mental states and environment. As expected, intuitions about these factors were affected by the operationalization of ensurance (i.e. manipulation). Specifically, whenever manipulation was present, participants were less inclined to agree that Jill was in control or that she had a choice about whether to steal.[12]

Second, one might object that verbal dispute theorists would make the same predictions, and thus the data from this experiment do not rule out their theory. After all, if "free will" is polysemous, we might see variation when only one factor is present. When both appear to be present or absent, there may be consensus about the presence and absence of free will, respectively. However, it isn't clear that the verbal dispute theory makes these predictions, even if we distinguish the following two versions: (1) the word "freely" is simply ambiguous (as with the word "bank") or (2) there are distinct populations with different dialects in which "freely" has a different meaning (as with the terms "coke" and "wicked" in parts of the U.S.).

Against the ambiguity version, consider what someone would say about a similar experiment on applications of the ambiguous word "bank." We would have to construct something like the following four scenarios about which we'd ask whether Sally "went to a bank." (a) Sally sits on the side of a river, then goes to deposit a check; (b) Sally sits at the park, then goes to deposit a check; (c) Sally sits on the side of a river, then goes to the grocery store; (d) Sally sits at the park, then goes to the grocery store. Presumably the ambiguity hypothesis is disjunctive in the sense that satisfying only one meaning of "bank" or the other should be sufficient for application of one of the concepts associated with the term. In that case, people should overwhelmingly agree that Sally went to a bank in all the scenarios except for the last one. If this is the model for "freely," then the predictions don't match the data, since the absence of liberty or ensurance alone significantly reduced ascriptions of freedom.

There is a different problem that is more pressing for the dialectical version, dealing with the fact that I randomly assigned participants to one of the vignettes. If we are dealing with different dialects, such random assignment should neutralize this difference and the two factors (liberty and ensurance) shouldn't have a systematic effect across the sample. After all, when a variable has a significant effect in experimental conditions involving random assignment, we have evidence that the effect occurs in the general population for the average person. So it is difficult to see how the data are compatible with there being distinct groups for whom "freely" expresses either a compatibilist or incompatibilist concept. Thus, the data are at least *prima facie* problematic for those positing a mere verbal dispute. Moreover, the same problems

[12] In ANOVAs where Control and Choice were now treated as dependent variables, there was a predicted main effect of Ensurance on intuitions about *control* [$F(1, 222) = 64.3$, $p < .001$, partial eta-squared = .224] as well as on intuitions about *choice* [$F(1, 222) = 39.8$, $p < .001$, partial eta-squared = .152].

would apply to those treating the terms "freely" (and its cognates) as having a different referent in different contexts (as in Hawthorne 2001).

4.2 Experiment 2: Avoiding Bypassing

In light of recent work done by Dylan Murray and Eddy Nahmias, one might worry about my description of a deterministic universe in which everything "must happen the exact same way." In two key experiments, Murray and Nahmias (forthcoming) provide support for the idea that one can generate seemingly incompatibilist intuitions in ordinary people only by getting them to misinterpret determinism. In particular, they suggest Nichols and Knobe's abstract scenario leads people to assume that the deterministic processes *bypass* the actor's psychology, yielding something more like fatalism than determinism (cf. also Feltz et al. 2009). Of course, if one's action is fated in that it will happen *no matter what*—regardless even of one's beliefs, desires, intentions, etc.—then ensurance is undermined. Yet saying everything in the universe *must* happen the same way, might encourage this misinterpretation of the deterministic universe. In that case, my participants' responses would be influenced, not by liberty and the lack of options determinism seems to engender, but instead by just another form of lacking the kind of control that is tied to ensurance.

To address this issue empirically and to attempt a replication of the previous results, I conducted a final experiment. I used the same design and materials as in the previous study, except a slight modification of the two vignettes in which Jill lacked liberty (Cells 3 and 4). I specifically replaced the two instances of "must" with "will"—both of which were in the first paragraph. The results, based on responses from 228 participants, are similar to those found previously (see Table 4).[13]

Table 4: Mean Freedom Responses for Experiment 2

	Ensurance	**No Ensurance**
Liberty	6.48 (SD=1.01)	3.80 (SD=2.21)
No Liberty	4.98 (SD=2.34)	2.92 (SD=2.09)

7-point scale with higher numbers corresponding to higher levels of agreement with the claim "In Universe 49, Jill stole the necklace freely." N = 228.

Not only are these means similar, an analysis of the data reveals that once again both ensurance and liberty had a statistically significant effect on responses.[14] However, while the differences are again statistically significant, the effect size of liberty was slightly lower and ensurance's slightly higher. This may indicate that the use of strong phrases such as "had to happen" or "must happen" (rather than "will happen") can have a slight

[13] I collected data from 241 participants, but responses from 13 were not included in the analysis, as they failed the comprehension check (same as in Experiment 1).

[14] I conducted a 2 (E vs. ~E) by 2 (L vs. ~L) between-subjects ANOVA. There was a main effect of Ensurance, $F(1, 224) = 80.9$, $p < .001$, partial eta-squared = .265, and a main effect of Liberty, $F(1, 224) = 20.36$, $p < .001$, partial eta-squared = .083. There was no interaction effect, $F(1, 224) = 1.412$, $p = .236$.

effect on responses, even though it does not fully account for the incompatibilist intuitions observed in previous studies.

I believe this should assuage the worry that the seemingly incompatibilist intuitions detected in the above studies are due merely to misinterpreting modal claims as involving bypassing or fatalism. On the contrary, lack of liberty due to determinism alone significantly reduces agreement with the claim that an agent acts freely.

However, without an explanation of why Murray and Nahmias were led to a different conclusion, the matter may seem unsettled. Moreover, Murray and Nahmias's project is to report evidence for an error theory for the incompatibilist intuitions of non-specialists. Since I propose that incompatibilist intuitions are more attuned to liberty and that this is part of our conceptual competence, my hypothesis is in trouble if their error theory is correct. So, in addition to their objection to Nichols and Knobe applying to the present experiment, their positive theory is a threat to the cluster hypothesis. Luckily, I believe there are several explanations in the offing for why Murray and Nahmias's results fail to establish their competing theory.

To see this, we need to understand their studies. The key strategy of Murray and Nahmias is to have subjects answer questions to determine whether they understood the scenario as involving determinism and not some sort of bypassing. Participants reported their degree of agreement or disagreement with statements such as "In Universe [A/C], what a person wants has no effect on what they end up doing. (What Bill wants has no effect on what he ends up doing.)" One should presumably *disagree* with such statements, even regarding a deterministic universe.[15]

In their first study, Murray and Nahmias sought to observe the relationship between bypassing and judgments about free will and moral responsibility. They collected responses to questions about free will, moral responsibility, and blame (yielding an *MR/FW composite score*), as well as responses to questions about bypassing (yielding a *bypassing composite score*). Subjects were randomly assigned to read and answer questions about one of four deterministic scenarios, which varied two factors: concreteness (abstract vs. concrete) and description type (Nichols and Knobe's vs. ones Nahmias has used with previous collaborators). The results suggested that, across the board, MR/FW scores tend to be lower when bypassing scores are higher (and vice versa). They also conducted a mediation analysis, which suggested that this was more than mere correlation: higher bypassing scores did seem to have some causal effect on lowering MR/FW scores.

In their second study, Murray and Nahmias attempted to explicitly control for bypassing. For two slightly modified deterministic vignettes, they added wording to explicitly avoid interpretations of the scenario as involving bypassing, such as:

> This does *not* mean that in Universe A people's mental states (their beliefs, desires, and decisions) have no effect on what they end up doing, and it does *not* mean that people are not part of the causal chains that lead to their actions. (§4.1)

[15] However, Joshua Shepherd (2012, p. 923) raises the interesting worry that some participants might claim a mental state "had no effect" on an agent's action only in the sense in which, say, a team's offense can have no effect on the opposition's defense. This is compatible with such factors playing a causal role.

Moreover, Murray and Nahmias wanted to make sure participants weren't mistaking what determinism *does* entail (in addition to misinterpreting what it doesn't). To this end, they checked for comprehension of the modal implications of determinism—namely, that "it *is* impossible, *holding fixed the past and the laws*, for future events to occur otherwise than they actually do" (§4). Their key prediction in this study was that participants who didn't conflate determinism with bypassing would by and large provide compatibilist intuitions. As with the first study, they began by excluding all the participants who failed the comprehension questions and found similar results. They then removed from analysis responses from those with a bypassing score at the midpoint or above. Most of those remaining reported compatibilist intuitions, in that the deterministic nature of the hypothetical universe did not significantly mitigate judgments of freedom or responsibility.

There are some substantial problems with their empirical argument. First, in both of their studies, Murray and Nahmias excluded responses from around half of their participants because they either "(a) responded incorrectly to either of two comprehension questions or (b) completed the survey quickly enough to indicate a lack of attention to the scenario and questions" (forthcoming, n. 22). In their crucial study intended to control for misunderstanding due to bypassing, they discarded *53%* of their subjects (161 of 302). This is a strikingly large portion of responses to leave out of the analysis. An initial worry is that, even if the comprehension questions seem unbiased, something seems awry from the outset. It is difficult to say what the comprehension questions were exactly. The authors provide two "sample comprehension questions" (in their Appendix). One example, which is much like the other, is: "According to the scenario, in Universe A, everything that happens is completely caused by what happened before it." The correct answer of course is "Yes." But one cannot be sure if some of the comprehension questions used were importantly different from the samples provided.

Second, the "correct" answers to the bypassing questions are not all uncontroversial. Most are, but one question is importantly suspect:

> **No Control:** In Universe [A/C], a person has no control over what they do. (Bill has no control over what he does.)

It is far from common ground that disagreeing with this statement to *some* degree is a misinterpretation of determinism. A deterministic universe surely does not yield that one's actions bypass all of one's mental states generally, but there is some room to argue that one is not clearly in control—and perhaps doesn't clearly make a full-blooded choice or decision—when there is only *one option*, given the past and the laws. And this may very well be implicit in some ordinary thinking. After all, people will apparently say our universe is more like an indeterministic one when presented with descriptions of both kinds—over 90% in Nichols and Knobe's study (2007, p. 669). So the ordinary conception of decision and control at least seems to arise in people who overwhelmingly believe the world is not deterministic.

Further support for this may come from Murray and Nahmias's own data purporting to show that determinism doesn't undermine judgments about an agent's abilities or having a choice. After reading the crucial abstract case (which seems to yield strong incompatibilist intuitions), only 31% of Murray and Nahmias's remaining subjects agreed that a person in that universe has "the ability to decide" to do something other

than what they actually decide to do (and only 58% for two of the other four scenarios). Without the ability to decide otherwise, one's control is presumably at least limited.

Murray and Nahmias's only explanation of this result is that this crucial vignette, originally from Nichols and Knobe, seems to be easily read as involving bypassing. But this is a stretch given that they have excluded the responses of more than half of their original set of participants, leaving a selected group who both passed the comprehension check and the bypassing questions. Moreover, this explanation fails to take seriously the idea that some participants may be assuming the agent in the deterministic universe has diminished control in a way that *doesn't* misinterpret determinism.[16]

In any event, there is a third and even more damaging problem, especially if it is combined with the others. While those who passed the comprehension check tended to provide compatibilist intuitions about a number of cases, *50%* of them had *incompatibilist* judgments about the abstract case modeled on Nichols and Knobe's (Murray & Nahmias forthcoming, n. 27). After further excluding any who provided answers indicating there was some bypassing in the case (or answered a related "modal question" incorrectly), only 62% provided apparently compatibilist responses about that abstract case. Yet these are the people who apparently did not misinterpret the deterministic universe as involving bypassing, according to their criteria. (The problem is worse if one of the bypassing questions is problematic.) So, even after controlling for a host of potential misinterpretations, the data from Murray and Nahmias show at the very least that certain scenarios can elicit genuinely incompatibilist responses from a sizable proportion of ordinary participants.

So, contrary to the bypassing theory, we don't need to attribute a performance error to ordinary people reporting incompatibilist intuitions. Rather, Murray and Nahmias's studies are consistent with the idea that both compatibilist and incompatibilist intuitions reside in ordinary thinking about freedom and responsibility, and the nature of the vignettes provided suggests that something like ensurance and liberty may play a role. Moreover, even if the bypassing theory were correct, it would only explain ordinary intuitions, not that of philosophers. After all, one wouldn't want to accuse the whole lot of specialists of making such an elementary confusion about a technical term like "determinism." So it cannot address the more general philosophical problem and the intense dialectical stalemate it engenders (compare Nichols 2006).

5. Conclusion

The cluster concept account is meant to help explain the long-standing debate about free will and determinism. It is undoubtedly bold to propose to do this, and I accordingly wish to tread lightly. Nevertheless, we arguably have good empirical and "armchair" reasons for the idea that the concept of free will is not associated with a single feature that is either compatible with determinism or not. Rather ensurance and liberty both play an important role in the concept of free will. When both are apparently present, one clearly

[16] Rose & Nichols (2013) have recently provided further empirical evidence that people are reluctant to grant the ability to decide in a deterministic universe. They argue that Murray and Nahmias have incorrectly claimed that bypassing judgments lead to incompatibilist intuitions when the causal order is actually the reverse.

acts in a way we would ordinarily describe as acting freely. If determinism is true with respect to our actions and we thus lack liberty, then free will is not obviously lost, and we do not need to say that we lack "libertarian free will" but retain the "compatibilist" counterpart. The same goes for indeterminism. Ironically enough, if determinism is true or false, it may simply be indeterminate whether we have free will, just as it is unclear whether some things are really games or furniture. But I submit this is an acceptable result given the mystery with which we began.

Some will remain tempted toward skepticism about our having free will; others may prefer to treat the debate as merely resting on a verbal dispute. I cannot do enough here to definitively rule out such alternatives, but I have suggested that each has substantial costs that my theory lacks. My main aim here is not to conclusively establish the truth of the cluster theory, but rather to sufficiently motivate it as a viable alternative to the classical accounts on offer. If we do have most reason to believe the cluster hypothesis, though, perhaps a productive attitude to take is one that is non-committal about whether free will requires or is precluded by determinism (cf. Mele 2006). A great deal of important philosophical work can still be done using this approach.

Acknowledgments: I owe a significant debt to John Maier who inspired me to test the main hypothesis in this paper, the basics of which we both independently developed. For valuable comments on or discussions about this paper, I thank: Lloyd Humberstone, Josh Knobe, Colin Marshall, Alfred Mele, Jonathan Phillips, Patrick Todd, Jason Turner, and the referees for this journal, one of whom kindly identified himself as Adam Feltz. Versions of the paper were presented at the University at Buffalo, Melbourne University, Deakin University, and CSU Wagga Wagga. Many thanks to the attendees for their feedback, especially Wylie Breckenridge, Daniel Cohen, Neil Levy, Edouard Machery, David Ripley, and Laura Schroeter. Work on this paper was supported by Monash University and some of the ideas were developed while participating in a summer seminar for the Big Questions in Free Will project at Florida State University, which was supported by the John Templeton Foundation. The views expressed in this article are solely the author's and do not reflect those of either funding body.

References

Balaguer, Mark (2010). *Free Will as an Open Scientific Problem*. MIT Press.

Buhrmester, M., T. Kwang, and S. D. Gosling (2011). "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?" *Perspectives on Psychological Science* 6.1: 3-5.

Chalmers, David J. (2011). "Verbal Disputes." *Philosophical Review* 120 (4):515-566.

Feltz, Adam (2013). "Pereboom and Premises: Asking the Right Questions in the Experimental Philosophy of Free Will." *Consciousness and Cognition* 22(1): 53–63.

Feltz, A., Cokely, E. T. & Nadelhoffer, T. (2009). "Natural Compatibilism versus Natural Incompatibilism: Back to the Drawing Board." *Mind and Language* 24 (1):1-23.

Feltz, A., Perez, A., & Harris, M. (2012). "Free Will, Causes, and Decisions: Individual Differences in Written Reports." *Journal of Consciousness Studies* 19(9-10): 166-189.

Fischer, J. M. (1994). *The Metaphysics of Free Will: An Essay on Control*. Malden: Blackwell.

Hawthorne, John (2001). "Freedom in Context." *Philosophical Studies* 104 (1): 63-79.

Knobe, Joshua and Doris, John (2010). "Responsibility." In *The Moral Psychology Handbook*, ed. John Doris. Oxford: Oxford University Press, pp. 321-54.

Knobe, Joshua (2014). "Free Will and the Scientific Vision." E. Machery and E. O'Neill (eds.), *Current Controversies in Experimental Philosophy*. Routledge.
Laurence, Stephen and Margolis, Eric (1999). "Concepts and Cognitive Science." In Laurence & Margolis (eds.) *Concepts: Core Readings*. MIT Press.
Maier, John (forthcoming). "The Agentive Modalities." *Philosophy and Phenomenological Research*.
May, Joshua & Holton, Richard (2012). "What in the World is Weakness of Will?" *Philosophical Studies* 157 (3):341–360.
Mele, Alfred (2006). *Free Will and Luck*. Oxford University Press.
Mele, Alfred (2014). "Free Will and Substance Dualism: The Real Scientific Threat to Free Will?" W. Sinnott-Armstrong, ed. *Moral Psychology, Volume 4: Free Will and Responsibility*, MIT Press.
Murray, Dylan & Nahmias, Eddy. (forthcoming). "Explaining Away Incompatibilist Intuitions." *Philosophy and Phenomenological Research*.
Nahmias, E., Morris, S. G., Nadelhoffer, T. & Turner, J. (2006). "Is Incompatibilism Intuitive?" *Philosophy and Phenomenological Research* 73 (1):28-53.
Nichols, Shaun (2006). "Folk Intuitions on Free Will." *Journal of Cognition and Culture* 6:57-86.
Nichols, Shaun & Knobe, Joshua (2007). "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions." *Noûs* 41 (4):663–685.
Park, John Jung (2013). "Prototypes, Exemplars, and Theoretical & Applied Ethics." *Neuroethics* 6 (2):237-247.
Rosch, Eleanor (1975). "Cognitive Representations of Semantic Categories." *Journal of Experimental Psychology: General* 104: 192–233.
Rose, David & Shaun Nichols (2013). "The Lesson of Bypassing." *Review of Philosophy and Psychology* 4(4): 599–619.
Sarkissian, H., Chatterjee, A., De Brigard, F., Knobe, J., Nichols, S. & Sirker, S. (2010). "Is Belief in Free Will a Cultural Universal?" *Mind and Language* 25 (3):346-358.
Shepherd, Joshua (2012). "Free Will and Consciousness: Experimental Studies." *Consciousness and Cognition* 21 (2):915-927.
Sinnott-Armstrong, Walter (2008). "Abstract + Concrete = Paradox." In J. Knobe & S. Nichols (eds.), *Experimental Philosophy*. Oxford University Press.
Sripada, Chandra S. (2012). "What Makes a Manipulated Agent Unfree?" *Philosophy and Phenomenological Research* 85(3): 563–593.
Strawson, P. F. (1962). "Freedom and Resentment." *Proceedings of the British Academy* 48, pp. 1-15.
van Inwagen, Peter (2000). "Free Will Remains a Mystery." *Philosophical Perspectives* 14:1-20.
van Inwagen, Peter (2008). "How to Think about the Problem of Free Will." *Journal of Ethics* 12 (3/4):327-341.
Woolfolk, R. L., Doris, J. M., & Darley, J. M. (2006/2008). "Identification, Situational Constraint, and Social Cognition: Studies in the Attribution of Moral Responsibility." In J. Knobe & S. Nichols (Eds.), *Experimental Philosophy*. Oxford: Oxford University Press, pp. 61-80.
Weigel, Chris (2011). "Distance, Anger, Freedom: An Account of the Role of Abstraction in Compatibilist and Incompatibilist Intuitions." *Philosophical Psychology* 24 (6):803-823.