What Do People Find Incompatible With Causal Determinism?

Adam Bear,[a] Joshua Knobe[b,c]

[a]*Department of Psychology, Yale University*
[b]*Program in Cognitive Science, Yale University*
[c]*Department of Philosophy, Yale University*

Abstract

Four studies explored people's judgments about whether particular types of behavior are compatible with determinism. Participants read a passage describing a deterministic universe, in which everything that happens is fully caused by whatever happened before it. They then assessed the degree to which different behaviors were possible in such a universe. Other participants evaluated the extent to which each of these behaviors had various features (e.g., requiring reasoning). We assessed the extent to which these features predicted judgments about whether the behaviors were possible in a deterministic universe. Experiments 1 and 2 found that people's judgments about whether a behavior was compatible with determinism were not predicted by their judgments about whether that behavior relies on physical processes in the brain and body, is uniquely human, is unpredictable, or involves reasoning. Experiment 3, however, found that a distinction between what we call "active" and "passive" behaviors can explain people's judgments. Experiment 4 extended these findings, showing that we can measure this distinction in several ways and that it is robustly predicted by two different cues. Taken together, these results suggest that people carve up mentally guided behavior into two distinct types—understanding one type to be compatible with determinism, but another type to be fundamentally incompatible with determinism.

Keywords: Free will; Determinism; Causation; Morality; Experimental philosophy

1. Introduction

One of the oldest and most venerable problems in philosophy is the problem of causal determinism. Suppose that everything that happens in the universe is caused in a com-

---

Correspondence should be sent to Adam Bear, Department of Psychology, Yale University, 2 Hillhouse Avenue, New Haven, CT 06511. E-mail: adam.bear@yale.edu

pletely deterministic way by some prior event. If the universe actually works in that way, can human beings still be free or morally responsible?

In recent years, a number of cognitive scientists have investigated people's ordinary intuitions about this question. When participants are presented with a vignette about a causally deterministic universe, they tend to say that the agents in this universe could not possibly be morally responsible for anything they do (Feltz & Millan, 2015; Nichols & Knobe, 2007; but see Murray & Nahmias, 2014). Moreover, this same tendency emerges in participants from a wide variety of different cultures (Sarkissian et al., 2010). These results suggest that people think that human actions would be different in some important respect if they were causally determined.

We extend this result in a new direction. Previous research showed that, in a determin- istic universe, people think that human beings cannot be morally responsible for their actions. We argue for an even stronger conclusion: In a deterministic universe, people think that certain kinds of actions would not even be possible at all. As we show below, people think that it would be possible in a deterministic universe to *compose complicated pieces of music*, but that it would not be possible to *think twice about doing something immoral*. Similarly, people think that in a deterministic universe, it would be possible for people to *feel depressed*, but that it would not be possible for people to *have the will- power to resist an urge to eat a steak in front of them*.

This result seems to point to something important about the way people understand human action. First, it enables us to pick out a class of actions that people understand in a quite distinctive way. Second, it shows that people think of actions within this class as deeply different from actions of other types, even to the point of having different intu- itions about the possible ways in which they can be produced.

1.1. Free will and determinism

The problem of free will has a long and celebrated history in philosophy. In particular, philosophers have argued over whether people can be morally responsible for their actions in a causally deterministic universe (e.g., Kant, 2005/1781; Nietzsche, 1954/1889; Spinoza, 2000/1667). This debate has raged on among philosophers and psychologists to the present day, with no clear end in sight (e.g., Baumeister, 2008; Greene & Cohen, 2004; Mele, 2006; Roskies, 2006). Given that our own universe may at least partially resemble a causally deterministic one, the answer to this question has obvious importance for how we understand ordinary human action and guide legal policy.

In recent years, though, psychologists and experimental philosophers have considered a similar but slightly different question, asking whether ordinary people *think* free will and moral responsibility are compatible with causal determinism. This research has uncovered many exciting results and has shown that people's intuitions about whether free will is possible in a deterministic universe are mixed and vary depending on various contextual and motivational factors (e.g., Clark, Luguri, Ditto, & Knobe, 2014; Deery, Davis, & Carey, 2015; Nadelhoffer, Shepard, Nahmias, Sripada, & Ross, 2014; Nichols, 2011; Nichols & Knobe, 2007; Paulhus & Carey, 2011; Pizarro & Helzer, 2010; see also

Cushman & Young, 2009). Moreover, belief in free will has been shown to influence morally relevant behavior like cheating (Vohs & Schooler, 2008), job performance (Stillman et al., 2010), and helping (Baumeister, Masicampo, & DeWall, 2009).

A small amount of existing research has been devoted to a question that might be regarded as even more basic. Independent of any issues about the relationship with free will and moral responsibility, do people think that our universe is causally deterministic to begin with? Research on this topic has used a number of different methodologies (Kushnir, Wellman, & Chernyak, 2009; Nichols, 2004), but one standard approach has been just to ask participants directly about the nature of our universe. Participants are given descriptions of two universes, a causally deterministic universe (Universe A) in which everything that occurs is completely caused by something that happened before it and a libertarian universe (Universe B) in which almost everything is caused by prior events but in which human decision making has a special status such that it need not be caused by anything at all (Nichols & Knobe, 2007). Studies using this method find that American participants tend to think our own universe is more like the libertarian Universe B (Nichols & Knobe, 2007), that participants in India, Hong Kong, and Colombia make the same judgment (Sarkissian et al., 2010), and that participants think key aspects of our present psychology could not even exist in the deterministic Universe A (Murray & Nahmias, 2014).

In work using this method, participants are given a vignette about a universe in which everything that happens is causally determined, that is, in which everything is completely caused in a deterministic way. To do this, participants are first given information about *causation*. (The universe is such that everything is caused by something that happened previously.) They are then given information about *determinism*. (The universe is such that it had to happen, given what occurred in the past, that everything in the present would happen exactly as it does.) These two pieces of information might seem highly abstract and divorced from ordinary experience. Nonetheless, we propose that people's intuitions about a universe that has these qualities can reveal something important about the way they understand the mind.

1.2. Dividing the mind

Consider the contrast between two cases:

1. An agent hears the loud bark of a dog and instinctively trembles.
2. An agent is tempted to perform some immoral action but then decides on reflection to do the right thing.

People might think an agent could perform the former action even if everything was causally determined, but would they think she could do the latter? What about a case in which the agent completes a series of complicated math problems? Or one in which the agent falls madly in love?

We propose that people carve up the mind in systematic ways, such that a certain kind of mental process is thought to be compatible with causal determinism, but another kind is thought to be fundamentally incompatible with causal determinism. Thus, a behavior like

instinctively trembling might be construed as falling in one category, while a behavior like resisting a temptation to do something immoral might be construed as falling in the other.

In testing this hypothesis, we ask what particular principles people are using to partition the mind into processes that are compatible or incompatible with causal determinism. Previous research unrelated to free will has uncovered a number of ways in which ordinary people divide up mental processes. In light of this research and other considerations, we examined five possible hypotheses about how people might distinguish between these categories:

1. *Mind–body dualism*: Psychological research suggests that people are inclined toward a form of mind–body dualism whereby they think some mental activities take place in the brain and body while others do not (Bloom, 2004, 2007). For example, people think that hunger and thirst are purely physical, but certain desires and beliefs are not (Bering & Bjorklund, 2004). Therefore, one intuitive hypothesis is that people not only think of brain and bodily activity as straightforwardly compatible with determinism, but they also believe that certain psychological processes depend on something that goes beyond the brain (e.g., an immaterial soul) and that these processes are incompatible with determinism.

2. *Human pursuits versus animalistic drives*: Along similar lines, people distinguish between "lower" animalistic activities and "higher," uniquely human, mental activities. For example, existing research suggests that temperamental traits, like neuroticism, are judged to be relatively animalistic, whereas traits relating to moral sensibility, civility, and the like are thought to be uniquely human (Haslam, 2006; Haslam, Bain, Douge, Lee, & Bastian, 2005). So another hypothesis is that it is easier to interpret more simple animal-like behavior in terms of deterministic causation, whereas uniquely human behaviors are not thought of in this way.

3. *Predictable versus unpredictable*: A third hypothesis it that people will be more inclined to see behaviors as not compatible with causal determinism when they are unpredictable. Specifically, if people could easily have predicted that something would happen (e.g., a person going to sleep after feeling tired), then they will be able to explain it in terms of a simple causal theory. But this might not be the case for more unpredictable behaviors (e.g., a person feeling love at first sight), and people might therefore conclude that these behaviors were not fully caused by prior states of the world (see Nichols, 2004, pp. 489–490).

4. *Reason versus instinct*: People distinguish between behaviors resulting from straightforward instinct or emotion and behaviors that require more sophisticated reasoning (Inbar, Cone, & Gilovich, 2010). The capacity for reasoning may then be seen as involving something mysterious or metaphysically distinctive. For example, Descartes (1946/1637) argued that straightforwardly instinctive behaviors could be generated in a purely mechanistic fashion, but that no merely mechanistic cause could mimic the operation of genuine reasoning. Perhaps ordinary people have a similar view, thinking that instinct and emotion are compatible with causal determinism but that reasoning is not.

5. *Active versus passive*: People seem to construe certain behaviors as passively "going with the flow," while construing other behaviors as actively "seizing control" and taking things in a different direction. For example, consider an agent who keeps going to the same boring job year after year. If this agent just stays in her job, people might feel that she is passively going with the flow, whereas if she impulsively quits her job and decides to become an artist, she might be seen as actively seizing control. Perhaps the processes people construe as "active" are precisely the ones that people think are incompatible with causal determinism because they are perceived as working counter to the normal causal forces in the world.

1.3. The present studies

In the experiments that follow, separate groups of participants rated a number of behaviors along the various dimensions that we hypothesized might relate to views about deterministic causation (dualism, unique humanness, unpredictability, reason vs. instinct, and activeness vs. passiveness). We then assessed the predictive power of judgments about our various hypotheses in predicting another set of participants' judgments about causal determinism on the basis of these participants' mean ratings for each of the behaviors.

To measure people's views about causal determinism, we adopted the standard approach of presenting participants with a vignette that describes a universe (Universe A) in which everything that happens is completely caused by something that happened before (Murray & Nahmias, 2014; Nichols & Knobe, 2007). Participants were then asked whether they agreed that various behaviors were possible in this universe (e.g., riding a bicycle or resisting the urge to eat a steak).

We measured each variable using a unique group of participants in order to reduce worries about demand characteristics, which have called similar work into question (see Scholl, 2008). Specifically, with this design, there was no way for participants to guess that we were looking, for example, at the relationship between judgments about causal determinism and judgments about dualism because no single participant would answer questions about both of these variables. Moreover, participants who we asked about causal determinism only received a question about one of our behaviors in order to ensure that they were thinking about the causal determinism prompt carefully and not assuming that we expected them to judge certain behaviors to be incompatible.

Over the course of four studies, we rule out the first four of the proposed hypotheses listed above, but find evidence for the last one that people's judgments about causal determinism follow a distinction between active versus passive behavior.

2. Experiment 1

In this experiment, some participants were asked to rate whether certain behaviors are possible in a causally deterministic universe and others rated these behaviors along

dimensions corresponding to our hypotheses. We focused on the dualism, human pursuits versus animalistic drives, and unpredictability hypotheses (tested, respectively, by the "Physical," "Human," and "Unexpected" items below). We also tested another prediction having to do with controlling impulses, which was consistent with both of the two remaining hypotheses.

2.1. Method

Our experiment consisted of two parts. First, we collected data on our dependent measure looking at how compatible our participants thought certain psychological behaviors are in a causally deterministic universe. Second, we collected data from separate groups of participants on "predictor" variables—variables we thought might explain which psychological behaviors were more or less possible in a deterministic universe.

2.1.1. Compatibility with causal determinism

Four hundred and fifty participants from Amazon's Mechanical Turk were assigned to make judgments on our dependent measure looking at how possible certain behaviors are in a causally deterministic universe (hereafter referred to as "Compatible"). To avoid demand characteristics and to encourage participants to think carefully about the question being asked, each participant judged only a single behavior. (In particular, we did not want participants to feel compelled to rate any behavior as incompatible with causal determinism if they did not believe this to be the case, which they might have been tempted to do if they were given multiple behaviors to rate.)

Participants were first presented with the following passage on causal determinism (adapted from Nichols & Knobe, 2007):

> 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 moved his leg. Like everything else, this action was completely caused by what happened before it. So, if everything in this universe was exactly the same up until John acted in this way, then it had to happen that John would move his leg.

After reading this passage and answering a comprehension question that followed ("Suppose that in Universe A, Sharon buys a television. Was this action completely caused by something that happened earlier?" [yes/no]), participants were randomly assigned to a question about one of the 30 possible behaviors that asked them to agree or disagree (on a 1–7 scale ranging from "strongly disagree" to "strongly agree") with a statement like "In Universe A, people can fall in love" (see Table 1 for full list of behaviors).

2.1.2. Predictor variables

One hundred and twenty new participants were assigned to one of four predictor conditions, yielding approximately 30 participants per predictor. The predictors were the following:

(Physical)   People do this thing purely as the result of physical processes in their brain and body.

Table 1
Mean ratings for Experiment 1 behaviors in ascending order of compatibility with determinism (DV)

Behavior	DV	E. Control	Unexpected	Physical	Human
Have the willpower to resist an urge to eat a steak in front of them	2.86	5.30	4.15	4.61	4.42
Suddenly change their mind about what they were planning to do	2.93	4.19	4.92	4.35	3.16
Think twice about doing something immoral	3.69	4.59	3.96	4.39	5.03
Decide to take the train instead of the bus to work	3.71	4.19	3.96	4.39	5.58
Perform actions that go against the values that their parents taught them	4.00	4.93	4.04	4.13	4.52
Contemplate deep philosophical problems	4.06	4.48	3.46	3.81	5.52
Have deep moral convictions	4.13	4.19	3.58	4.26	4.97
Evaluate the evidence for and against a view	4.14	5.19	3.88	4.23	5.55
Have a planned schedule for their day	4.15	4.96	3.31	4.10	4.71
Try their hardest to win a race	4.20	4.37	4.04	5.03	3.29
Worry about paying their bills on time	4.27	4.30	3.96	4.16	6.00
Carry on complex conversations	4.40	4.56	4.04	4.42	4.71
Unconsciously resent their bosses	4.87	3.78	4.58	4.13	4.87
Have moral values	4.93	4.30	4.15	4.13	5.16
Want to go to the grocery store and therefore drive there	5.00	3.85	3.96	4.65	5.32
Believe that it is raining and, as a result, grab an umbrella before walking outside	5.07	3.37	3.88	4.74	5.10
Construct complex legal arguments	5.13	4.70	2.96	4.10	5.61
Compute answers to simple math problems like 2 + 3 = 5	5.13	4.33	4.19	4.94	4.48
Get angry and stomp their foot on the floor as a result	5.13	3.37	4.15	5.06	2.87
Learn new skills, like how to play the piano	5.13	4.74	3.54	4.84	4.29
Find pleasure in the beauty of a sunset	5.18	3.15	4.96	4.65	4.13
Compose complicated pieces of music	5.40	4.07	3.50	4.35	5.55
Imagine a purple square	5.44	3.89	3.50	4.29	4.48
Feel depressed	5.45	2.89	3.85	4.32	2.81
Reflexively release their hands from a hot stove	5.50	2.67	4.92	5.29	2.94
Know facts like the fact that there are seven continents in the world	5.60	4.04	4.15	4.48	5.87
Become thirsty in the middle of the night and, because of this thirst, sleepwalk to the fridge	5.63	3.11	4.35	4.71	3.90
Have cravings for ice cream that cause them to buy this food	5.81	3.30	5.08	4.81	4.68
Fall in love	6.06	3.22	5.04	4.52	3.97
Ride bicycles	6.13	3.78	4.19	4.84	4.39

(Human)   People can only do this thing because of special properties of human minds that distinguish them from other animals.

(Unexpected)   People do this thing unexpectedly and unpredictably.

(Exert Control) People can only do this thing by exerting control over their instinctual behavior.

Each participant who was assigned to one of these predictors rated the assigned predictor statement on a 1–7 scale ("strongly disagree" to "strongly agree") for all 30 of our behaviors, presented in random order. Two attention checks were also randomly interspersed within this list (asking participants to answer "Strongly Agree" and "Somewhat Agree" for those list items).

2.2. Results

Fourteen participants who failed our comprehension check (for the determinism dependent measure) and seven participants who failed at least one of the attention checks (for the predictor conditions) were excluded from analysis. The analyses that follow were done at the level of the behaviors themselves, collapsing across ratings from individual participants. In other words, each behavior had a unique mean rating for each variable (see Table 1), and these mean ratings were treated as predictive characteristics in subsequent analyses.

We first assessed correlations between our individual predictors and our dependent measure Compatible. Of the four predictors tested, only two—Physical and Exert Control—were significantly correlated with this dependent measure, $r(28) = .37$, $p = .047$ and $r(28) = -.67$, $p < .001$, respectively. That is, behaviors associated with a physical brain or body were more likely to be judged to be possible in a causally deterministic universe, and behaviors involving more inhibitory control were judged to be less possible in a causally deterministic universe. Similar correlations for the predictors Unexpected, $r(28) = .17$, $p = .36$, and Human, $r(28) = -.12$, $p = .53$, did not reach significance.

We next tested the predictive power of our four predictors in a regression model with Compatible as our dependent measure, which was found to be highly significant, $F(4, 25) = 8.58$, $p < .001$, $R^2 = .58$. In this analysis, only Exert Control was found to be a significant predictor, $b = -1.03$, $SE = .21$, $\beta = -.84$, $p < .001$, but none of the other three predictors were (all ps > .05).

2.3. Discussion

There was substantial variability in the behaviors that people rated as compatible with causal determinism, with some behaviors rated as far more compatible (e.g., falling in love, $M = 6.06$) than other behaviors (e.g., resisting the urge to eat a steak, $M = 2.86$). This fact is, in itself, somewhat surprising. As discussed, philosophers have long

questioned whether people can be morally responsible for their actions in a causally deterministic universe, but our results suggest that ordinary people think that individuals could not even *perform* certain actions in such a universe, regardless of their moral content. Thus, people's intuitions about causal determinism may cut much deeper than just the moral domain.

We explored three plausible hypotheses about what might be driving differences in compatibility judgments across behaviors, but we failed to find evidence for them. Specifically, we failed to find support for the ideas that (a) people think that behaviors arising from nonphysical processes (e.g., an immaterial soul) are less possible in a causally deterministic universe, (b) people think that uniquely human behaviors are less possible in a causally deterministic universe, and (c) people think that unpredictable behaviors are less possible in a causally deterministic universe. Of course, though these results suggest that these hypotheses may be incorrect, they could also have arisen for a number of reasons that have nothing to do with the incorrectness of the hypotheses (limited statistical power, the specific behaviors used, etc.). Thus, more work is needed to definitively show that factors like dualistic cognition play no role in judgments about causal determinism.

Whatever the source of these null results, we did find strong support for the idea that exerting control over instinctual behavior is crucial to what people find incompatible with causal determinism. We explore this idea further in Experiment 2.

3. Experiment 2

Experiment 2 focused on carefully testing the reason versus instinct hypothesis. This hypothesis predicts that behaviors involving more reason should be judged to be incompatible with causal determinism, whereas more emotional or instinctual behaviors should be judged to be compatible. This might explain why in Experiment 1 Exert Control was such a strong predictor of what people found to be incompatible with causal determinism: People may think that willpower, or exerting control over instincts, requires the capacity for reasoning, which is in turn incompatible with causal determinism. (Of course, there may instead be other features of willpower besides reason that people find to be at odds with causal determinism, as explored in Experiment 4.)

We measured ascriptions of reason using two new predictor variables. One of these predictors simply asked whether a given behavior requires reasoning, while another asked specifically about cases in which reason "wins out" against emotion or instinct. To test this latter hypothesis, we needed to ensure that any possible effect of the "Wins Out" predictor on determinism judgments was not simply explained by people thinking that behaviors are more likely to be incompatible with determinism when they involve any kind of psychological conflict. Therefore, we included a more generic "Conflict" predictor to compare to the "Wins Out" predictor and to serve as a control. In addition, we designed new behaviors in order to distinguish between conflicts in which an agent uses reason to overcome instinct or emotion (e.g., deciding not to go to a party in order to

study for an exam) from conflicts in which an agent ends up taking the action that is impulsive or erratic (e.g., suddenly quitting a stable, but boring, job in order to become an artist).

3.1. Method

Experimental procedures for Experiment 2 were identical to those of Experiment 1, except that we used 20 new behaviors and three new predictors designed to more carefully test our new hypotheses. Three hundred participants from Mechanical Turk were randomly assigned to answer the same question about possibility in a causally deterministic universe ("Compatible") from Experiment 1 for one of 20 of these behaviors, and a different 90 participants were randomly assigned to rate all 20 behaviors on one of the following predictor statements:

(Conflict)   When people do this thing, they experience a conflict within themselves.

(Reason)   In order to do this thing, people need to use the reasoning or reflective parts of their minds.

(Wins Out)   When people do this thing, their reason or rationality wins out against their emotions or impulses.

As in Experiment 1, these 20 behaviors were presented in random order, along with two attention checks.

3.2. Results

Ten participants who failed our comprehension check (for the determinism condition) and nine participants who failed at least one of our attention checks (for the predictor conditions) were excluded from further analysis. As in Experiment 1, analyses were done at the level of the behaviors themselves, collapsing across participants' individual responses. Means for these behaviors on Compatible and our three new predictors are presented in Table 2.

Of our three new predictors, only Conflict was significantly correlated with Compatible, with behaviors involving more psychological conflict being judged as less compatible with causal determinism, $r(18) = -.48$, $p = .030$. Reason and Wins Out, however, had no significant relationship with Compatible, $r(18) = .21$, $p = .38$ and $r(18) = .15$, $p = .52$, respectively.

In a regression model predicting Compatible based on all three of these predictors, no predictor was significantly predictive at the $p < .05$ level, and these predictors did not explain a statistically significant amount of variance in this dependent measure, $F(3, 16) = 1.84$, $p = .18$, $R^2 = .26$. Nevertheless, Conflict had a marginally significant effect, $b = -.46$, $SE = .22$, $\beta = -.46$, $p = .053$.

Table 2
Mean ratings for Experiment 2 behaviors in ascending order of compatibility with determinism (DV)

Behavior	DV	Conflict	Reason	Wins Out
Resist the urge to eat a steak	2.64	5.20	4.35	4.50
Impulsively shoplift from a convenience store	2.73	4.60	2.35	2.67
Leave a comfortable home on a whim	3.40	4.73	3.08	3.00
Leave their long-term spouse unexpectedly to pursue a new love interest	3.69	5.43	2.88	2.88
Think twice about doing something immoral	4.07	5.63	5.46	5.58
Decide not to go to a party in order to study for an exam	4.20	5.10	5.46	5.71
Improvise a line of music that doesn't follow the rules of music theory that they were taught	4.20	4.47	4.85	3.42
Splurge on a new television	4.27	4.20	3.12	3.04
Remain in a difficult marriage for the good of their children	4.50	6.00	4.88	4.92
Have complicated political discussions	4.63	4.30	5.54	4.88
Suddenly quit a stable, but boring, job in order to become an artist	5.08	5.03	4.19	2.92
Try to remember somebody's phone number without writing it down	5.13	3.60	5.15	5.08
Compose a highly technical legal document	5.20	3.07	5.81	5.38
Solve a complicated math problem	5.27	3.27	5.81	5.83
Design a careful science experiment	5.29	2.97	6.00	5.88
Satisfy strong sexual desires	5.53	3.73	3.50	3.54
Commit suicide despite living seemingly happy lives	5.73	5.60	2.69	2.29
Honor a promise they made to someone a long time ago even though they don't really want to	5.80	5.47	4.92	5.46
Dance to catchy music at a nightclub	6.00	2.70	3.00	3.38
Play sports	6.36	2.87	4.19	3.67

3.3. Discussion

Experiment 2 found that ascriptions of reason did not significantly predict compatibility judgments. This point is clearly exemplified by the two behaviors rated most incompatible with causal determinism. One of these behaviors seems to be a paradigm case of reason winning out over emotion or instinct (resisting an urge to eat a steak), but the other seems to be a paradigm case of precisely the opposite sort of phenomenon (impulsively shoplifting from a convenience store). Of course, again, the lack of an observed relationship between reason and compatibility judgments could be explained by lack of statistical power or any number of other factors.

Surprisingly, the results showed a marginal effect for the Conflict predictor, which was not intended to test any of our main hypotheses. We further explore this effect in the next study.

4. Experiment 3

Experiment 2 tentatively ruled out our reason/instinct hypothesis but suggested that some other feature of psychological conflict could be accounting for variation in causal

determinism judgments. Experiment 3 was designed to test whether our active/passive hypothesis could be playing this role.

To understand the key distinction here at an intuitive level, imagine a conflict within the mind of a heroin addict. On one hand, he might want to keep using heroin because of the pleasure he gets from it, but on the other hand, he might want to stop using heroin because of its dangers. Intuitively, there is an important asymmetry between the two paths this agent can take. One path (using heroin) might be thought of as "passive." It might be seen as letting oneself be controlled by one's environment, "going with the flow" or living one's life "like clockwork." The other option (resisting heroin) might be thought of as "active." It might be seen as taking initiative, "breaking off from the flow of things" or "seizing control" and taking things in a different direction.

In this one example, the "active" path is also the one that people might associate with careful reasoning (as opposed to emotion or instinct), but there are other cases in which these two notions seem to come apart. Imagine an agent who goes to a convenience store to get some food. If she buys the food in the normal way at the register, we might see her behavior as passive ("going with the flow," "like clockwork"), whereas if she suddenly and impulsively decides to shoplift the food, we might see her behavior as active ("breaking off from the flow of things" and going in a different direction). Here the path that we might intuitively see as active is not the one we would most associate with careful reasoning.

Of course, a question immediately arises about what criteria people use in drawing the active/passive distinction. We will be addressing that question below (in Experiment 4 and in the General Discussion), but for the moment we put it aside. Instead, we simply test the hypothesis that the active/passive distinction—however exactly it is drawn—predicts people's intuitions about causal determinism.

To test this hypothesis, we designed a new predictor condition in which participants read a description of the kind of active behavior we hypothesized might be related to beliefs about causal determinism and compared participants' ratings of our behaviors in this condition to responses on the more general Conflict predictor from Experiment 2. We also included some new behaviors, in addition to those from Experiment 2, which we thought would be construed as high in conflict, but "passive" (e.g., slowing down at the end of a close race from fatigue). As such, these new behaviors would help disambiguate the active/passive hypothesis from a more general view that people think that *any* behavior involving psychological conflict is incompatible with causal determinism. On the active/passive hypothesis, these high-conflict, but passive, behaviors should be judged to be compatible.

4.1. Method

The same basic design from Experiments 1 and 2 was again implemented in this experiment, with two changes. First, we randomly assigned 450 participants to one of 30 behaviors, 20 of which were the same as those from Experiment 2 and 10 of which were

new and designed to be relatively high in conflict but low on our new Initiative measure (see Table 3 for full list of behaviors). Second, we only tested two predictor variables (30 participants each): Conflict from Experiment 2 and a new predictor measuring the amount of initiative people need to take to perform our behaviors and go against the "various forces in their minds and environments." Because this new Initiative predictor was measuring a more complex concept, we provided these slightly longer directions to participants assigned to this condition:

> Sometimes people just let themselves be controlled by the various forces in their minds and environments. At other times, they need to take initiative and work against these forces.

Table 3
Mean ratings for Experiment 3 behaviors in ascending order of compatibility with determinism (DV)

Behavior	DV	Conflict	Initiative
Resist the urge to eat a steak	2.46	4.09	4.76
Leave a comfortable home on a whim	2.56	4.88	4.64
Improvise a line of music that doesn't follow the rules of music theory that they were taught	2.69	3.78	4.52
Impulsively shoplift from a convenience store	2.93	4.47	4.28
Decide not to go to a party in order to study for an exam	3.13	4.88	5.04
Think twice about doing something immoral	3.33	5.72	4.76
Splurge on a new television	3.67	4.78	4.20
Continue to date their attractive, but uncaring, partner	4.27	5.22	4.20
Have complicated political discussions	4.33	4.53	4.52
Design a careful science experiment	4.47	3.31	4.20
Spend money at an expensive restaurant that they considered giving to charity	4.60	5.03	3.96
Suddenly quit a stable, but boring, job in order to become an artist	4.87	5.09	4.92
Remain in a difficult marriage for the good of their children	4.87	5.63	5.28
Order a chocolate cake for dessert instead of something healthier	4.92	4.59	3.68
Let the government oppress them without protesting or demanding justice	4.93	5.28	3.92
Honor a promise they made to someone a long time ago even though they don't really want to	4.93	5.19	5.00
Try to remember somebody's phone number without writing it down	5.06	3.31	4.48
Let a bully harass their close friend out of fear of getting bullied themselves	5.07	5.41	3.60
Satisfy strong sexual desires	5.07	4.03	4.40
Keep the same unpleasant, but lucrative, job year after year	5.08	5.44	4.56
Stay at a fun party instead of doing important work	5.13	5.00	4.36
Spend the day relaxing at home instead of going on a more exciting day trip	5.42	3.53	3.60
Slow down at the end of a close race from fatigue	5.47	4.53	3.16
Commit suicide despite living seemingly happy lives	5.53	5.50	4.96
Leave their long-term spouse unexpectedly to pursue a new love interest	5.75	5.13	4.48
Play sports	5.81	2.84	3.76
Continue to use heroin despite efforts to stop	5.93	5.69	3.84
Solve a complicated math problem	5.93	3.44	4.04
Dance to catchy music at a nightclub	6.00	2.59	3.56
Compose a highly technical legal document	6.00	3.09	4.04

For each of the following behaviors, we want to know whether you agree that people can only do it by taking the kind of initiative that involves working against various forces in their minds and environments.

As in the other experiments, two attention checks were randomly included along with the list of behaviors.

4.2. Results

Three participants who failed our attention checks (for the predictor conditions) and 16 who failed our comprehension check (for the determinism condition) were excluded from further analyses. Once again, all analyses were done at the level of the behaviors.

Unlike in Experiment 2, Conflict was not significantly correlated with Compatible, $r(28) = -.18$, $p = .33$. Our new Initiative predictor, however, was significantly negatively correlated with this dependent measure, with greater initiative being associated with lower compatibility with causal determinism as hypothesized, $r(28) = -.44$, $p = .016$. In a regression predicting Compatible that included both of these predictors, the overall equation was marginal, $F(2, 27) = 3.20$, $p = .057$, $R^2 = .19$. Only Initiative had a significant (negative) relationship, $b = -.90$, $SE = .39$, $\beta = -.43$, $p = .030$, whereas Conflict did not, $b = -.02$, $SE = .22$, $\beta = -.02$, $p = .93$.

4.3. Discussion

The active/passive hypothesis seemed to give the best account of what people judge to be impossible in a causally deterministic universe. Crucially, conflict was not in itself a cue to what people found incompatible with causal determinism, but only a particular kind of action taken when presented with a conflict was. For example, continuing to use heroin despite efforts to stop was judged not only to be high in conflict, but also compatible with causal determinism because it was a "passive" behavior for which somebody is pushed around by the various forces in their mind and environment. Conversely, "active" behaviors performed in the face of conflict, like deciding not to go to a party in order to study for an examination, were definitively judged to be incompatible with determinism.

But what kind of factors influence whether a behavior is characterized as active or passive? Experiment 4 sought to make these folk concepts more precise by identifying particular cues that lead people to classify a behavior as active or passive, which, in turn, influence judgments about causal determinism.

5. Experiment 4

Experiment 3 provided tentative evidence that the active/passive distinction predicts people's causal determinism judgments. In this final experiment, we try to obtain a better

understanding of how people conceptualize this distinction and what cues they use to determine whether an agent is behaving actively or passively.

First, we included two additional measures of the active/passive distinction, which invoked different metaphors that measure the same general construct. Together, the three measures should enable us to assess active/passive judgments in a way that is not sensitive to the idiosyncrasies of any one specific presentation or description. More important, the existence of three differently described, but highly correlated, measures of this presumed distinction would provide compelling evidence that this distinction is a genuine psychological construct.

Second, we measured two potential cues people may use to determine whether a behavior was performed actively or passively. The results of our previous experiments suggested that people tend to regard behaviors as incompatible with causal determinism both when these behaviors require a great deal of mental *effort* (e.g., resisting the urge to eat a steak) and when they are performed on a whim, erratically, or *spontaneously* (e.g., impulsively shoplifting). Specifically, with regard to mental effort, Experiment 1 suggested that exerting control over one's instincts was a strong predictor of what people found incompatible with causal determinism, but Experiment 2 suggested that reason, on its own, did not predict these incompatibility judgments. So people seem to regard exerting control as incompatible with causal determinism not because it involves reason, but because it requires something else, such as mental effort more generally. With regard to spontaneity, Experiment 2 suggested that certain behaviors that seem spontaneous are judged to be incompatible with determinism, but we did not formally test this hypothesis.

Although effort and spontaneity are fundamentally opposing qualities in certain respects, we hypothesized that they would both serve as cues that a behavior was performed actively rather than passively. In addition, we hypothesized that if these cues had any impact on causal determinism judgments, this impact would be mediated by people's active/passive judgments.

5.1. Method

The basic procedure from previous experiments was again implemented here. However, we did not re-collect data for Compatible, but instead combined the responses on this measure for all behaviors from Experiment 3 and most of the behaviors from Experiment 1 (yielding 54 behaviors in total). A few of Experiment 1's behaviors were not included in the study because they were too similar to behaviors from Experiment 3 (e.g., "have the willpower to resist an urge to eat a steak in front of them").

One hundred and fifty participants were randomly assigned to one of the five conditions. Three of these conditions measured the active/passive distinction. The first of these (referred to as "Initiative" in Results) was identical to our description from Experiment 3. The other two were the following:

(Flow)   Sometimes people just passively go with the flow, while at other times, they seem to be actively breaking off to do something different.

For each of the following behaviors, we want to know whether you agree that people can only do it by actively breaking off from the flow of things to do something different.

(Clockwork)   Sometimes people's actions seem to proceed mechanically like clockwork. Other times, people seize control and take things in a different direction.
                   For each of the following behaviors, we want to know whether you agree that people can only do it by seizing control in this way.

Two separate groups of participants were asked to rate the 54 behaviors based on descriptions of our two predicted cues for the active/passive distinction:

(Effort)   When people do this thing, they need to exert mental effort to overcome some desire they have.

(Spontaneity)   When people do this thing, they do it spontaneously, erratically, or on a whim.

As before, two attention checks were randomly included in this list of behaviors in all conditions.

5.2. Results

Twenty-two participants were eliminated from analysis because they failed at least one attention check. Once again, all analyses were conducted using the mean ratings for each of the behaviors on our tested variables. Mean ratings for the dependent measure of possibility in a deterministic universe were taken from Experiments 1 (Table 1) and 3 (Table 3).

First, we assessed the internal consistency of our three active/passive measures (Initiative, Flow, and Clockwork). These measures were found to have good consistency (Cronbach's α = .84) and were therefore averaged into a single active/passive measure for correlational analysis. Consistent with our findings from Experiment 3, this averaged active/passive measure was found to be significantly negatively correlated with Compatible, $r(52) = -.46$, $p < .001$, with more active behaviors being judged to be less possible in a causally deterministic universe. In addition, in a regression model regressing the active/passive composite on Effort, $b = .63$, $SE = .07$, $\beta = .92$, $p < .001$, and Spontaneity, $b = .26$, $SE = .06$, $\beta = .44$, $p < .001$, both of these cues were found to significantly predict active/passive judgments, $F(2, 51) = 39.53$, $p < .001$, $R^2 = .61$. (Because these two cues were negatively correlated with each other, $r(52) = -.54$, $p < .001$, however, only Effort, $r(52) = .68$, $p < .001$, but not Spontaneity, $r(52) = -.05$, $p = .70$, showed a zero-order correlation with active/passive judgments.)

To assess how these variables relate to judgments about causal determinism, we ran a series of regression analyses (see Table 4). In all cases, the dependent measure was Compatible. First, we conducted an analysis with only the two cues of Effort and Spontaneity as predictors. Both cues significantly predicted Compatible (ps < .05), $F(2, 51) = 3.84$,

$p = .028$, $R^2 = .13$. Next, in addition to including these two predictors, we added their interaction in a further regression model, which was marginally predictive, $F(3, 50) = 2.75$, $p = .052$, $R^2 = .14$. Both Effort and Spontaneity remained significant predictors of Compatible ($p < .05$), but their interaction was nonsignificant ($p > .1$). Finally, we included the active/passive composite in a further regression model, which was highly significant, $F(4, 49) = 3.90$, $p = .008$, $R^2 = .24$. The active/passive composite was found to significantly predict Compatible ($p < .05$), and importantly, Effort and Spontaneity became nonsignificant ($ps > .1$).[1]

To investigate the relationship between Effort, Spontaneity, the active/passive composite, and Compatible more carefully, we conducted two bootstrapped mediation analyses (Preacher & Hayes, 2008). First, we examined whether the effect of Effort on Compatible was mediated by the active/passive distinction when Spontaneity was entered as a covariate (Fig. 1). This analysis revealed that the influence of Effort on causal determinism judgments was, in fact, mediated by active/passive judgments (95% bias-corrected CI = −1.15 to −.08). We then conducted a similar analysis examining whether Spontaneity's influence on Compatible was mediated by the active/passive distinction when Effort was entered as a covariate (Fig. 2). This analysis also revealed significant mediation (95% bias-corrected CI = −.51 to −.03). In short, the impact of both Effort and Spontaneity on causal determinism judgments was partially mediated by the active/passive distinction.

5.3. Discussion

Experiment 4 provides more compelling evidence that a distinction between active and passive behaviors predicts what behaviors are viewed as deterministically caused. In addi-

Table 4
Regressions on judged compatibility with determinism from Experiment 4

Predictor	Compatible (1)	Compatible (2)	Compatible (3)
Effort	−.41* (.19)	−.43** (.20)	.03 (.30)
Spontaneity	−.32* (.16)	−.32* (.16)	−.10 (.18)
Effort × Spontaneity		−.11 (.21)	−.13 (.20)
Active/passive			−.51* (.37)

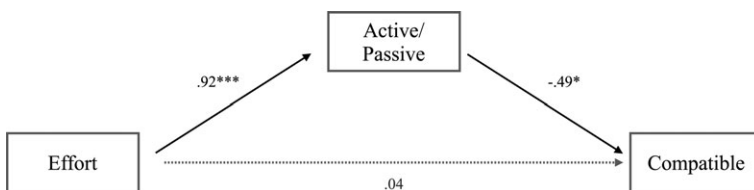*Note.* All coefficients standardized, with standard errors in parentheses. **$p < .01$, *$p < .05$.

Fig. 1. Mediation analysis from Experiment 4, with Spontaneity as a covariate. All coefficients reported are standardized values. Statistically significant paths are denoted by solid lines and nonsignificant paths by dotted lines. ***$p < .001$, *$p < .05$.
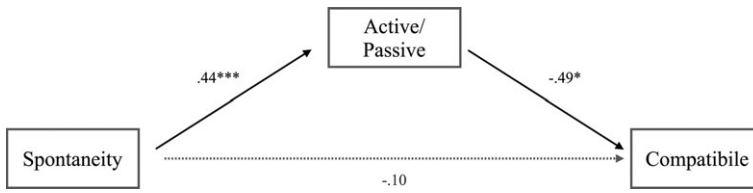
Fig. 2. Mediation analysis from Experiment 4, with Effort as a covariate. All coefficients reported are standardized values. Statistically significant paths are denoted by solid lines and nonsignificant paths by dotted lines. ***$p < .001$, *$p < .05$.

tion, it overcomes the shortcomings of Experiment 3, confirming that the active/passive distinction is a real psychological construct that can be measured in several different ways and that is systemically driven by cues relating to mental effort and spontaneity. Importantly, this distinction was found to predict judgments about causal determinism even when controlling for the interaction between mental effort and spontaneity—suggesting that the distinction plays an important role in explaining causal determinism judgments over and above the role of just the cues.

It is worth noting that these two cues themselves were significantly negatively related to each other, suggesting that they are conceptually quite distinct. The kind of behaviors that require mental effort (e.g., resisting an urge to eat a steak) differ from the kind of behaviors that involve spontaneity (e.g., leaving a comfortable home on a whim) both in terms of the psychological processes involved and the character of their phenomenology. Although the former behaviors involve higher reasoning faculties and feel very effortful, the latter behaviors often involve more emotional processes and generally have far less of a subjective experience of effort. Consequently, the active/passive distinction cannot be straightforwardly captured by either simple psychological distinctions or by appealing to phenomenological experience. Instead, it seems to be driven by more complex theoretical principles. We discuss this issue further in the General Discussion.

6. General discussion

Four studies examined people's judgments about causal determinism. These studies consistently showed that certain behaviors are judged to be, in and of themselves, incompatible with causal determinism. Experiment 1 suggested that distinctions related to dualism, uniquely human mental states, or unpredictability could not fully explain what behaviors were thought of in this way. Experiment 2, further, suggested that a distinction between reason and emotion could not account for our pattern of results. Experiment 3 provided evidence that behaving "actively," in particular, is a concept people find irreconcilable with deterministic causation. Finally, Experiment 4 further supported this conclusion by testing a model in which two cues, involving mental effort and spontaneity, influence judgments about what behaviors are active or passive, which in turn explain judgments about determinism.

The use of a fully between-participant design ensured that our results were not due to any conscious theorizing about the nature of the experiment. Participants who evaluated our question about causal determinism had no exposure to our other questions about, for example, dualism or unpredictability, and therefore would not be able to easily infer our study's purpose. Moreover, because each participant who received the determinism question was only asked about one behavior, they could not deduce anything about the range of behaviors we were asking about and would not have any incentive to answer our determinism question in a patterned way.

As discussed in the Introduction, the most basic result found in all of our studies—that people judge some behaviors as inherently incompatible with causal determinism—suggests something important about the way people conceptualize human action. People may not only think that moral responsibility requires the ability to exercise some kind of causally indeterministic "free will," but also think that certain behaviors themselves must be produced in a causally indeterministic way and could not be performed otherwise.

In addition to making this general point about how people reason about action, our findings are of interest for three more specific reasons. First, they tell us that certain distinctions may not actually be relevant to people's views about determinism and causality despite having a plausible connection. Second, they introduce a novel distinction between active and passive behaviors, which ought to be explored further. Third, they connect up in interesting ways with the existing literature in the experimental philosophy of free will. These topics are discussed below.

6.1. Failed predictors of determinism

In conjunction with our positive findings, it is surprising and interesting to learn that certain plausible hypotheses about what could explain people's judgments about determinism were not supported in the present studies. Of course, these null results could be explained by a number of factors that might not generalize to future work—the small sample of behaviors used, the particular behaviors chosen, the phrasing of the predictor statements, and so on. But it is nevertheless striking that we found no evidence of an effect of these predictors at all.

Logically speaking, it makes sense that a distinction between mental processes that occur in a material brain versus an immaterial soul might explain why people think certain behaviors could or could not happen if everything were causally determined. Brain activity seems like the kind of process that could be interpreted in straightforward causal terms (e.g., a brain state in motor cortex causes certain muscles to move, which produces a certain behavior), so things attributed to the brain should also be deemed possible in a deterministic universe. Conversely, the workings of an immaterial soul might be less readily interpreted in this causal way because souls differ from ordinary physical entities for which we observe causation (e.g., one billiard ball hitting another). Surprisingly, though, emerging evidence suggests that views about determinism and dualism are largely uncorrelated (Nadelhoffer et al., 2014) and that judgments about what kind of beings have free will do not depend on whether or not these beings have a soul (Monroe, Dillon,

& Malle, 2014; see also Monroe & Malle, 2010). Thus, our findings provide further support for the view that beliefs about determinism importantly differ from beliefs about dualism.

It was additionally notable to find that unpredictable or unexpected behaviors were not judged to be more incompatible with determinism than predictable behaviors. As Nichols (2004, pp. 489–490) points out, people may be inclined to explain behavior in indeterministic terms when predictions from their ordinary causal theories of the world are violated or generally inadequate. But our results suggest that at least a strong version of this thesis is false: Certain behaviors that were judged to be fairly predictable, like resisting the urge to eat a steak, were judged to be incompatible with determinism, and other behaviors that were judged to be unpredictable, like falling in love, were judged to be compatible with determinism. This latter finding provides empirical support for Nichols' (2004) criticism of the unpredictability view—namely, that many highly unpredictable phenomena, like the weather, seem to work in completely causally deterministic ways.

Finally, it was surprising to find that neither more cognitively complex nor more uniquely human behaviors were judged to be less compatible with determinism than simpler or more animalistic behaviors. At first blush, behaviors like solving a complicated math problem or falling in love are precisely the kind of behaviors whose workings are mysterious and indecipherable in a way that might make them difficult to encompass within a normal causal theory of how the mind works. Yet it is clear from our data that this kind of mystery is not the source of people's intuitions about determinism.

It is true that cognitively complex behaviors that specifically involve cognitive control or willpower (e.g., resisting the urge to eat a steak) were judged far less compatible with determinism than most other behaviors. Crucially, though, spontaneous behaviors, which mostly involve the *absence of* cognitive control, were also highly predictive of people's determinism judgments. And, as we show in Experiment 4, these two kinds of behavior only predict determinism judgments in so far as they are cues to the active/passive distinction. In other words, it is not cognitive control itself that explains people's intuitions about determinism, but a broader theoretical construct of which cognitive control is only one part.

6.2. Active versus passive behavior

Although there are many reasons why variables like dualism might not have significantly predicted judgments about causal determinism in our studies, we found compelling evidence that a distinction between "active" and "passive" behaviors could largely explain these judgments. To ensure that this distinction was a genuine psychological construct, we measured it using three distinct phrasings (being "controlled by forces" vs. taking "initiative," "going with the flow" vs. "breaking off from the flow of things," proceeding mechanically "like clockwork" versus "seizing control" and taking things in a different direction) and found that it was robustly predicted by two concrete cues involving mental effort and spontaneity.

This finding illuminates what we believe to be a striking connection between a folk psychological notion concerning what is active versus passive and a question about the metaphysical structure of the universe. Although people's commonsense ideas about taking initiative, going with the flow, and the like seem to be straightforwardly psychological descriptions of human behavior, it was found that these judgments actually predict beliefs about what kinds of things are possible in a causally deterministic universe. In light of the surprising connection between these two seemingly unrelated ideas, our results call for a more nuanced examination of how people conceptualize this active/passive distinction.

In Experiment 3, participants characterized passive behavior as being "controlled by various forces in one's mind or environment." One natural hypothesis, then, is that people have an implicit theory about how these various psychological forces work and the ways they might lead a person to act "passively." This theory might be similar in many ways to the theories people use to make sense of purely inanimate physical objects, but it would also differ from those theories in one important respect. People's theories about purely physical objects dictate that these objects will actually move in the way the forces act upon them. In other words, if people knew all the forces acting on a billiard ball, they could predict exactly where it was going to move. In contrast, people's theories of psychology might not dictate that people must do what the psychological forces are pushing them to do. People might instead think that these forces in an agent's mind are drawing her to behave in a particular way, but they would not think that she necessarily has to behave that way. More specifically, they might think that forces are drawing her to act in a "passive" manner, but that she also has the capacity to "actively" resist these forces and do something else instead.

On this hypothesis, people's understanding of active behavior would be radically different from their understanding of passive behavior. Their understanding of passive behavior would be much like their understanding of purely physical events. By contrast, people's understanding of active behavior could not be properly characterized in this way. Rather, people might have a theory about how agents act when behaving passively, and they construe an agent's behavior as active simply whenever she does something other than what this theory predicts. For example, suppose that people's implicit theory tells them that the forces within an agent's mind are drawing her to keep working at the same job day after day. If she does continue working at that job, her behavior will be seen as passive, whereas if she somehow manages to diverge from that path (through mental effort, on a spontaneous impulse, or in any other way), her behavior will be seen as active.

In short, it may be that there is nothing in particular in common among active behaviors. Different active behaviors are merely judged to be active in virtue of *not* being passive. Indeed, this might explain why, in our experiments, effort and spontaneity were both found to be cues to what is active despite being deeply different, and even opposing, qualities.

6.3. Limitations

Along with raising new questions, the present studies suffer from several limitations. For one thing, although we have argued that cues about mental effort and spontaneity

explain judgments about causal determinism via the active/passive distinction, further work is needed to build a more nuanced picture of the relationship among these variables. For example, our studies did not consider how people reason when somebody exerts effort to overcome a morally *good* urge to do something morally bad (we only considered the opposite). Moreover, the present studies did not examine behaviors in which mental effort was exerted to perform an act that might be considered spontaneous. How might people understand a case in which, for instance, a workaholic finally overcomes her drive to keep working in order to take a relaxing vacation?

More generally, we have not provided an explanation for why people would draw a distinction between active and passive behaviors in the first place. One possible hypothesis would be that this distinction is tapping into a genuine feature of the way the human mind works. That is, it might be thought that a proper scientific theory of the mind would include something at least relatively similar to the active/passive distinction (see, e.g., Cooper, Ruh, & Mareschal, 2014), and it might then be suggested that people come to draw this distinction ordinarily because they are tracking, with a least some degree of accuracy, a fact about the workings of the mind. A second possible hypothesis would be that the function of this distinction is actually best understood not as a matter of tracking scientific facts about the mind but rather of doing something else entirely. (For example, it might be thought that the primary function of this distinction is in making judgments of moral responsibility.) Finally, a third hypothesis would be that the active/passive distinction does not have a function at all and is instead simply a byproduct of more basic processes people use to reason about other minds or about the world. The present studies cannot decide between these competing hypotheses.

6.4. Free will and moral responsibility

As discussed in the Introduction, research in the field of experimental philosophy has focused on the question of whether ordinary people think that moral responsibility is compatible with causal determinism. This research has indicated that, under certain circumstances, people think that agents could not be morally responsible in a deterministic universe (Nichols & Knobe, 2007). However, there is considerable debate about how to explain this result. Some researchers have suggested that these findings demonstrate something fundamental about how people reason about the mind (e.g., Knobe, 2014; Nichols & Knobe, 2007; Rose & Nichols, 2013), while others have suggested that the results might merely show that people fail to correctly understand what determinism is about (e.g., Murray & Nahmias, 2014; Nahmias & Thompson, 2014).

In the present studies, participants did not receive questions about moral responsibility, so the results do not bear directly on this question. However, the results may bear on issues of moral responsibility in a more indirect way, by pointing to a new hypothesis that could be explored in further work. Specifically, the present results suggest that people think "passive" behavior is possible in a deterministic universe, but "active" behavior is not. It may be precisely this aspect of people's cognition that makes them so reluctant to regard agents in a deterministic universe as morally responsible.

For example, suppose that people are considering an agent in a deterministic universe who is controlled by her fears and passively allows her friend to be bullied. The present results reveal something surprising about the way people understand behaviors like this one. They suggest that people think it actually would not be possible for any agent who was in this sort of situation in a deterministic universe to actively work against her fears and take a different path to confront the bully. Perhaps it is this fact about people's judgments that makes them reluctant to say that agents in a deterministic universe can be morally responsible. Future work measuring moral judgment more directly could help assess whether people are, in fact, reasoning in this way.

7. Conclusion

Over the course of four experiments, we found that people think certain types of behaviors are fundamentally incompatible with causal determinism. We tested five hypotheses about what kinds of behaviors might be thought of in this way. In the end, only one of these hypotheses, having to do with active versus passive behaviors, explained a significant portion of the variance in people's beliefs. Despite its use in ordinary discourse, this distinction has not been studied scientifically until this point. However, our results suggest that it ought to be examined more closely, as it may provide deep and important insights into how people reason about free will, morality, and the mind more generally.

Acknowledgments

Note

1. An anonymous reviewer points out that the results for our Compatible measure might in part have arisen because participants think that these behaviors would not be possible even in the actual universe. To explore this hypothesis, we recruited an additional sample of participants ($N = 810$), randomly assigning each of them to one of the behaviors and asking them to rate (on a scale from 1 to 7) whether it would be possible to perform that behavior in our actual universe. When we enter this measure of possibility in the actual universe into the regression equation along with our other predictors, there is indeed a marginally significant effect of

possibility in the actual universe (β = .23, p = .070), but active/passive remains significant (β = $-$.54, p = .009).

References

Baumeister, R. F. (2008). Free will in scientific psychology. *Perspectives on Psychological Science*, *3*(1), 14–19.

Baumeister, R. F., Masicampo, E. J., & DeWall, C. N. (2009). Prosocial benefits of feeling free: Disbelief in free will increases aggression and reduces helpfulness. *Personality and Social Psychology Bulletin*, *35*(2), 260–268.

Bering, J. M., & Bjorklund, D. F. (2004). The natural emergence of reasoning about the afterlife as a developmental regularity. *Developmental Psychology*, *40*(2), 217.

Bloom, P. (2004). *Descartes' baby: How the science of child development explains what makes us human*. New York: Basic Books.

Bloom, P. (2007). Religion is natural. *Developmental Science*, *10*(1), 147–151.

Clark, C. J., Luguri, J. B., Ditto, P. H., & Knobe, J. (2014). Free to punish: A motivated account of free will belief. *Journal of Personality and Social Psychology*, *106*, 501–513.

Cooper, R. P., Ruh, N., & Mareschal, D. (2014). The goal circuit model: A hierarchical multi-route model of the acquisition and control of routine sequential action in humans. *Cognitive Science*, *38*(2), 244–274.

Cushman, F., & Young, L. (2009). The psychology of dilemmas and the philosophy of morality. *Ethical Theory and Moral Practice*, *12*(1), 9–24.

Deery, O., Davis, T., & Carey, J. (2015). The free-will intuitions scale and the question of natural compatibilism. *Philosophical Psychology*, *28*(6), 776–801.

Descartes, R. (1946). *Discourse on the method of rightly conducting the reason, and seeking truth in the science*. La Salle, IL: Open Court. (Original work published 1637.)

Feltz, A., & Millan, M. (2015). An error theory for compatibilist intuitions. *Philosophical Psychology*, *28*(4), 529–555.

Greene, J., & Cohen, J. (2004). For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *359*(1451), 1775–1785.

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review*, *10*(3), 252–264.

Haslam, N., Bain, P., Douge, L., Lee, M., & Bastian, B. (2005). More human than you: Attributing humanness to self and others. *Journal of Personality and Social Psychology*, *89*(6), 937.

Inbar, Y., Cone, J., & Gilovich, T. (2010). People's intuitions about intuitive insight and intuitive choice. *Journal of Personality and Social Psychology*, *99*(2), 232.

Kant, I. (2005). *The critique of pure reason* (P. Guyer & A.W. Wood, Trans.). New York: Cambridge University Press. (Original work published 1781.)

Knobe, J. (2014). Free will and the scientific vision. In E. Machery & E. O'Neill (eds.), *Current controversies in experimental philosophy* (pp. 69–85). New York: Routledge

Kushnir, T., Wellman, H. M., & Chernyak, N. (2009). Preschoolers' understanding of freedom of choice. *Proceedings of the Thirty-First Annual Meeting of the Cognitive Science Society* 87–92.

Mele, A. R. (2006). *Free will and luck*. New York: Oxford University Press.

Monroe, A. E., Dillon, K. D., & Malle, B. F. (2014). Bringing free will down to earth: People's psychological concept of free will and its role in moral judgment. *Consciousness and Cognition*, *27*, 100–108.

Monroe, A. E., & Malle, B. F. (2010). From uncaused will to conscious choice: The need to study, not speculate about people's folk concept of free will. *Review of Philosophy and Psychology*, *1*(2), 211–224.

Murray, D., & Nahmias, E. (2014). Explaining away incompatibilist intuitions. *Philosophy and Phenomenological Research*, *88*, 434–467.

Nadelhoffer, T., Shepard, J., Nahmias, E., Sripada, C., & Ross, L. T. (2014). The free will inventory: Measuring beliefs about agency and responsibility. *Consciousness and Cognition*, *25*(C), 27–41.

Nahmias, E., & Thompson, M. (2014). A naturalistic vision of free will. In E. Machery & E. O'Neill (Eds.), *Current controversies in experimental philosophy* (pp. 86–103). New York, NY: Routledge.

Nichols, S. (2004). The folk psychology of free will: Fits and starts. *Mind & Language*, *19*(5), 473–502.

Nichols, S. (2011). Experimental philosophy and the problem of free will. *Science*, *331*(6023), 1401–1403.

Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous*, *41*(4), 663–685.

Nietzsche, F. (1954). *Twilight of the idols* (W. Kaufmann, Trans.). New York: Penguin Books. (Original work published 1889.)

Paulhus, D. L., & Carey, J. M. (2011). The FAD–Plus: Measuring lay beliefs regarding free will and related constructs. *Journal of Personality Assessment*, *93*(1), 96–104.

Pizarro, D. A., & Helzer, E. (2010). Freedom of the will and stubborn moralism. In R. Baumeister, K. Vohs, & A. Mele (Eds.), *Free will and consciousness: How might they work?* (pp. 101–120). Oxford, England: Oxford University Press.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891.

Rose, D., & Nichols, S. (2013). The lesson of bypassing. *Review of Philosophy and Psychology*, *4*(4), 599–619.

Roskies, A. (2006). Neuroscientific challenges to free will and responsibility. *Trends in Cognitive Sciences*, *10*(9), 419–423.

Sarkissian, H., Chatterjee, A., De Brigard, F., Knobe, J., Nichols, S., & Sirker, S. (2010). Is belief in free will a cultural universal? *Mind & Language*, *25*(3), 346–358.

Scholl, B. J. (2008). Two kinds of experimental philosophy (and their methodological dangers). In *SPP Workshop on Experimental Philosophy*. Philadelphia, PA: University of Pennsylvania.

Spinoza, B. (2000). *Ethics* (G. Parkinson, Ed.). Oxford, England: Oxford University Press. (Original work published 1667).

Stillman, T. F., Baumeister, R. F., Vohs, K. D., Lambert, N. M., Fincham, F. D., & Brewer, L. E. (2010). Personal philosophy and personnel achievement: Belief in free will predicts better job performance. *Social Psychological and Personality Science*, *1*(1), 43–50.

Vohs, K. D., & Schooler, J. W. (2008). The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychological Science*, *19*(1), 49–54.