The Lesson of Bypassing[1]

[*forthcoming* in Review of Philosophy and Psychology]

David Rose
Department of Philosophy
Rutgers University

Shaun Nichols
Department of Philosophy
University of Arizona

1. Intuitions, Incompatibilism, and Bypassing

The idea that incompatibilism is intuitive is one of the key motivations for incompatibilism. Not surprisingly, then, philosophers who defend incompatibilism often claim that incompatibilism is the natural, commonsense view about free will and moral responsibility (e.g., Pereboom 2001, Kane 1999, Strawson 1986). This claim has received some support from empirical studies that indicate that when given a description of a determinist universe, participants tend to say that a person in that universe could not be morally responsible for his actions (Nichols & Knobe 2007; Roskies & Nichols 2008; Sarkissian et al. 2010). There are, of course, sophisticated arguments for incompatibilism (e.g. Pereboom 2001; van Inwagen 1983), but the claim that incompatibilism resonates with commonsense is an important consideration in its favor.

In a series of important papers, Eddy Nahmias and colleagues have argued that judgments that seems to reflect a lay commitment to incompatibilism are really driven by a confusion (Nahmias et al. 2007; Nahmias & Murray 2011; Murray and Nahmias forthcoming). When given the description of determinism, they suggest, people mistakenly interpret determinism to mean that our mental states lack causal efficacy. Nahmias and colleagues dub this "bypassing". In their recent paper, Nahmias and Dylan Murray (Murray & Nahmias forthcoming) characterize the basic idea of bypassing as follows: "An agent's mental states are bypassed when she ends up doing what she does regardless of what she had thought, wanted, or tried to do." (MS, pp. 5-6). If people's mental states are bypassed this is a very good reason to think that people don't have free will or moral responsibility. However, determinism does not entail bypassing. After all, even if determinism is true, our mental states might well be critical causal factors in generating our behavior. So if people think that determinism entails bypassing, this would explain why people respond in ways that conform to incompatibilism; but it wouldn't show that people are really incompatibilists, since it's a confusion to think that determinism entails bypassing. As we'll see in more detail shortly, Nahmias and colleagues have found evidence that people do indeed often seem to think that determinism implies bypassing.

[1] We are grateful to David Danks, Joshua Knobe, Dylan Murray, Eddy Nahmias, and three anonymous referees for helpful comments on a draft of this paper.

The bypassing hypothesis is that the allegedly incompatibilist judgments are really mediated by a confusion – the confusion of drawing a bypassing inference from a description of determinism. The idea is ingenious, plausible, and widely admired (e.g. Holton 2009; Mele forthcoming, Sommers 2012). There are now several studies showing that people think that mental states are bypassed if determinism is true (Knobe forthcoming; Nahmias et al. 2007; Nahmias & Murray 2011; Murray & Nahmias forthcoming).

In this paper, we will argue that the bypassing results have been systematically misinterpreted. The fact that people give bypassing responses to deterministic scenarios does reveal something important about how people think about deterministic scenarios. But far from showing that incompatibilist responses are borne of a confusion, we will argue that the bypassing results suggest that incompatibilism runs even deeper than previously recognized.

2. Bypassing and causal models

Although there is evidence in favor of the bypassing hypotheses, not all is rosy for the bypassing theory. First, people tend *not* to think that determinism implies bypassing for natural processes like volcanoes (Deery et al. forthcoming). When determinism is explained to subjects, they will say that even if determinism is true, whether a volcano occurs depends on the physical processes that precede the volcano. So the confusion about determinism is not global – people don't end up confusing determinism with a global fatalism such that volcanoes are simply fated to occur when they do, regardless of causal antecedents. In addition, Joshua Knobe found that people do not make bypassing judgments about computers. He described a deterministic universe to subjects, asked them to imagine that the universe includes "computers that use programs and data", and then asked them to indicate whether they agree that in that universe "the computers' programs and data have no effect on what they end up being caused to do". He found that people tended to *disagree* with that statement, suggesting that they do *not* think computer programs are bypassed. In a subsequent task, he asked instead about facial expressions. Participants were asked whether they agreed that in the deterministic universe, "people's emotions have no effect on what facial expressions they end up making". Again, people tended to disagree with this statement.

All this generates something of a puzzle. If people really do confuse determinism with bypassing, why is the confusion so narrowly focused? Why don't people make the confusion about physical events, computers, or facial expressions? Or, perhaps the better question is why *do* people make the bypassing judgments about mental states and action?

2.1. Two Causal Models

One possibility, which is suggested by Murray and Nahmias, is that the description of determinism leads people to infer a specific form of bypassing – the bypassing of mental states in the generation of behavior. From this (mistaken) inference to bypassing, people then draw the judgment that there is no moral responsibility or free will. That is, according to the model, when participants read a description of a Determinist universe, this leads them to infer that the production of behavior in that universe Bypasses mental states, and this affects their responses to questions about moral responsibility and free will (MR/FW) – in particular, it leads to judgments

that people in that universe lack moral responsibility and free will. The resulting causal model is the *Bypassing Model*:

(i) Determinism→Bypassing→MR/FW

An alternative causal model holds that when participants read a description of a deterministic universe this leads them leads them to make incompatibilist judgments on the questions about moral responsibility and free will (MR/FW), and these judgments then lead to a judgment about the causal effects of mental states being bypassed in the generation of behavior. We can call this causal model the *MR/FW Incompatibilist Model*:

(ii) Determinism →MR/FW→Bypassing

Obviously the Bypassing and MR/FW Incompatibilist Models are *competing* causal models of the relationship between Determinism, Bypassing and MR/FW. But how are we to decide between these two models?

2.2. Support for the Bypassing Model

Murray and Nahmias use a statistical technique known as "mediation analysis" to argue in favor of the Bypassing Model. In their study, participants were assigned to one of two conditions:

Nichols and Knobe (2007) Abstract:[2]

> 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.
>
> Now imagine a universe (Universe B) in which almost everything that happens is completely caused by whatever happened before it. The one exception is human decision making. For example, one day Mary decided to have French Fries at lunch. Since a person's decision in this universe is not completely caused by what happened before it, even if everything in the universe was exactly the same up until Mary made her decision, it did not have to happen that Mary would decide to have French Fries. She could have decided to have something different.
>
> The key difference, then, is that in Universe A every decision is completely caused by what happened before the decision – given the past, each decision has

[2] "Abstract" here is to be contrasted with "Concrete". In "Concrete" cases a specific individual engaging in a particular behavior is described while in "Abstract" cases no particular individual or behavior is described. Some concrete cases tend to elicit more compatibilist judgments (see Nahmias et al. 2007; Nichols and Knobe, 2007).

> to happen the way that it does. By contrast, in Universe B, decisions are not completely caused by the past, and each human decision does not have to happen the way that it does.

Nahmias, Morris, Naddelhoffer and Turner (2006) Abstract:

> Imagine there is a universe (Universe C) 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 cause the exact same events for the entire history of the universe, so that every single time the universe is re-created, everything must happen the exact same way. For instance, in this universe whenever a person decides to do something, every time the universe is re-created, that person decides to do the same thing at that time and then does it.

In one condition, participants read only the Nichols and Knobe Abstract description; in the other condition, participants read only the Nahmias, Morris, Naddelhoffer and Turner (2006) Abstract description. These two descriptions composed the Determinism variable.

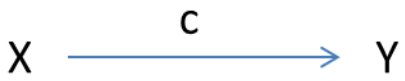
After reading one of the two descriptions, participants were asked to indicate their level of agreement, on a 6-pt. scale, to a series of questions—about either universe A (Nichols & Knobe) or C (Nahmias et al.) —which were then combined to create the MR/FW and Bypassing variables (see Table 1).

Variable Type	**Variable Name**	**Questions**
MR/FW	MR	In Universe [A/C], it is possible for a person to be fully morally responsible for their actions.
	FW	In Universe [A/C], it is possible for a person to have free will.
	Blame	In Universe [A/C], a person deserves to be blamed for the bad things they do.
Bypassing	Decisions	In Universe [A/C], a person's decisions have no effect on what they end up doing.
	Wants	In Universe [A/C], what a person wants has no effect on what they end up doing.
	Believes	In Universe [A/C], what a person believes has no effect on what they end up doing.
	No control	In Universe [A/C], a person has no control over what they do.

Table 1: MR/FW and Bypassing Questions

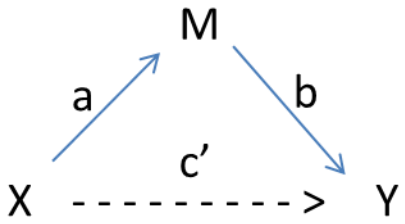
Murray and Nahmias followed a standard procedure—outlined in Baron and Kenny (1986)—for testing mediation. Since mediation plays such an important role in their argument, we want to explain the procedure in a bit of detail.

Let's begin by considering a model whereby a predictor variable X is assumed to cause an outcome variable Y:



In this model, path c is an estimation of the total effect that X has on Y. To estimate the total effect of X on Y one conducts a regression analysis in order to determine whether X is a significant predictor of the outcome variable Y. This is Step 1 in testing for mediation.

Now, in mediation analyses, the issue is whether some other variable, M, mediates the effects of X on Y. If M does mediate the effects of X on Y, we can represent it as follows:



There are a few things to note about this model. The first is that path *a* is an estimation of the effects of X on M. To test this, one conducts a regression analysis to determine whether X is a significant predictor of M. This is Step 2 in testing for mediation. Secondly, path *b* is an estimation of the effects of M on Y. Here again, one tests this by conducting a regression analysis. This is Step 3 in testing for mediation. Next, notice that path *c'* is dotted in this model. What this is essentially indicating is that there is no direct effect of X on Y, given M. That is, M mediates the effects of X on Y and so X only has an indirect effect on Y. In testing this, one conducts a multiple regression model using both X and M as predictors of Y, examining whether the total effect—*c*—of X on Y is eliminated when M is included in the model. Put another way, once the contribution of M on Y is controlled for, one examines whether X is still a significant predictor of Y. This is Step 4 in testing mediation.

It is important to note that if Steps 1-3 are not satisfied then one cannot infer mediation. In other words, in conducting regression analyses for Steps 1-3 if one finds that the predictor variable is not a significant predictor of the outcome variable, then one cannot go onto Step 4 to test for mediation. A final important point to note about mediation is that it assumes that the proposed mediator M causes the outcome variable Y, not the other way around. Often these causal assumptions are based on prior theoretical results that motivate the assumption, but, as will be of some importance below, it is critical to keep in mind that the direction of the causal arrow is just assumed in any *single* mediation analysis.

Returning now to Murray and Nahmias, they found that Steps 1-4 were satisfied such that the effects of Determinism on MR/FW were mediated by Bypassing. Thus, they take their results to provide support for the Bypassing Model.

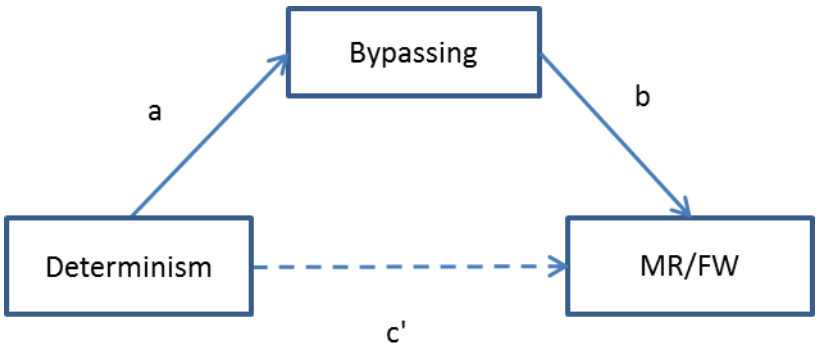


Figure 1: Mediation analysis

Based on these results, Murray and Nahmias suggest that "the difference in MR/FW responses between the two abstract conditions of the scenarios was largely *caused* by people's Bypassing judgments." (Murray & Nahmias forthcoming, MS p. 12)

2.3. Issues with the Bypassing Model

The mediation analysis offered by Murray and Nahmias provides some support for the Bypassing Model. But it does not provide *unique* support for the Bypassing Model. To see why, we need to add just a little more to our explanation of mediation analysis. Recall that a mediation model *assumes* that the mediator M causes the outcome variable Y and not the other way around. The Bypassing Model assumes that the causal direction goes from Bypassing to MR/FW. But this is a substantive assumption, and the fact that Bypassing mediates MR/FW in Figure 1 doesn't entail that MR/FW doesn't also mediate Bypassing. This is because "mediation" is really just a series of correlational measures. In other words, the regression analyses involved in testing mediation are essentially just correlations that assume a causal direction. Some of the causal assumptions may be perfectly legitimate while others are not. For example, in Figure 1, Determinism is the variable that is externally manipulated and so we know that Determinism cannot be caused by any other variables. It can only cause the values on other variables if it causes those values at all. But Bypassing and MR/FW are not externally manipulated, and so it is an open question what the causal relationship between Bypassing and MR/FW is. The mediation model that Murray and Nahmias offer tells us that if we know the Bypassing score, we can predict the MR/FW score pretty well. But that doesn't exclude the possibility that the flip side also holds. We might also be able to predict the Bypassing score if we know the MR/FW score. It's a familiar point about correlation. If you know the road is wet, you can predict that it's been raining; if you know it's been raining you can predict that the road is wet. But obviously the causal direction only holds for one of these predictive relationships. So the fact that they found that Bypassing mediates MR/FW doesn't conclusively show that Bypassing causes MR/FW any more than the fact that the variable WET-ROADS would mediate the variable RAIN shows that wet roads cause rain (see also Chan, forthcoming).

To be clear, we are *not* claiming that mediation does not provide *any* evidence of directionality. Rather, we are claiming that finding mediation for a *single* model does not provide conclusive evidence of directionality. The reason for this is that sometimes it is the case that the data provide support for two mediation models. In other words, sometimes the data can turn out in such a way that one finds *both* that M mediates the effects of X on Y and that Y mediates the

effects of X on M.[3] In such a case, one *cannot* infer directionality. Indeed, although it is not reported in their paper, Murray and Nahmias have now tested the alternative mediation model, and they find mediation for the MR/FW Incompatibilist model as well (personal communication). Since they find mediation for two alternative causal models (the Bypassing model and the MR/FW Incompatibilist model), it is *indeterminate* which causal model is the correct causal model of the data.

3. Study 1: Causal Models

Since the mediation analysis offered by Murray and Nahmias is both theoretically and empirically consistent with the MR/FW Incompatibilist model, we decided to conduct a new study in an attempt to decide between the Bypassing and MR/FW Incompatibilist Models.

Our experiment departed in two significant ways from Murray and Nahmias. First, we tried to minimize differences between the two conditions. Instead of asking about the two different descriptions of determinism, we used the same description of determinism (Nichols & Knobe, 2007) and varied whether we asked about a deterministic universe or an indeterminist one. Second, we constructed two structural equation models to test the alternative causal models. While running a series of regressions can often be useful for testing mediation, structural equation models are more discriminating, offering the advantage of providing a measure of *overall* fit for a model. Indeed, Iacobucci, Saldanha and Deng (2007) suggest that testing mediation models using structural equation models is always superior to running a series of regressions. Since the regressions used by Murray and Nahmias do not enable us to decide between the two alternative causal models, structural equation models plausibly offer an advantage in terms of discriminating between the models.

3.1. Design

All participants were given the Nichols and Knobe (2007) Abstract case (see Section 2.1) and then asked: "Which of these universes do you think is most like ours?" and given the option of selecting A or B. Then participants were asked to explain their answers.

We followed Murray and Nahmias and asked both MR/FW and Bypassing questions. For each question, participants were asked to indicate the extent to which they agreed or disagreed on a 6-point scale anchored at 1= "strongly disagree" and 6= "strongly agree" (See Table 1).[4] What we varied was whether these questions referred to Universe A or B.

Finally, we included two comprehension checks asking participants to indicate whether they thought the statements were true or false:

---

[3] One way this can happen is when M and Y are highly correlated. In Murray and Nahmias' study 1 (MS, p. 11), for example, they report a very high correlation between Bypassing and MR/FW, r=-.734.

[4] Following both Nichols and Knobe (2007) and Murray and Nahmias (forthcoming), the moral responsibility question was always presented first. The presentation of all the other questions were randomized to control for order effects.

(1) According to the scenario, in Universe A, nothing that happens is completely caused by what happened before it.

(2) According to the scenario, there is no difference between Universe A and Universe B.

3.2. Results

A total of 66 participants were recruited through Amazon's Mechanical Turk. Four of these participants missed one of the comprehension questions, and two failed to complete the survey, leaving 60 participants for the final sample.

We began by creating two composite scores—Bypassing and MR/FW—by combining responses for each item relevant to Bypassing and each item relevant to MR/FW (see Table 1).[5] Then, we constructed two separate structural equation models to test the Bypassing and MR/FW Incompatibilist Models:

Bypassing Model: Determinism→Bypassing→MR/FW

MR/FW Incompatibilist Model: Determinism→MR/FW→Bypassing

We began by conducting a causal search over our data with Tetrad IV,[6] using the Greedy Equivalence Search (GES) algorithm.[7] The model returned from GES was the MR/FW Incompatibilist Model.[8] The model indicates that Determinism directly causes MR/FW and that MR/FW causes Bypassing. The model also tells us that once we condition on MR/FW, the effects of Determinism on Bypassing are screened off. That is, once we factor in how people respond on MR/FW, Determinism does *not* predict Bypassing responses (see Figure 2).

[5] The average inter-correlations between the four items composing Bypassing (Cronbach's Alpha=.936) and the average inter-correlations between the three items composing MR/FW (Cronbach's Alpha=.936) were quite high, thereby indicating a high degree of internal consistency among the items composing each variable type.

[6] http://www.phil.cmu.edu/projects/tetrad/

[7] Essentially, GES considers the possible models available given the different variables. Each variable is assigned to a node, and the nodes are used to build the different possible models. GES begins by assigning an information score to the null model in which the nodes are all disconnected. GES then considers various possible arrows ("edges") between the different nodes. The algorithm will add the edge that yields the greatest improvement in the information score (if there is such an edge). The algorithm repeats the process, adding the next edge that makes for the greatest improvement in the information score. When the algorithm reaches the point where no new edges improve the information score, it proceeds to consider deleting edges. It first finds the edge for which deleting that edge would yield the greatest improvement in the information score (if there is such an edge). It repeats this procedure until no further deletions will improve the score. In all cases, the orientation of the edges is given by edge-orientation rules (Meek, 1997). It has been shown (Chickering, 2002) that, given enough data, GES will return the true causal model of the data. GES is often interpreted as returning the best fitting causal model, given the data. (For further details, see Chickering (2002) and Rose et al. (2011).)

[8] This model was a very good fit of the data, χ^2 (1)=.9026, p=.3421, BIC=-3.1918.

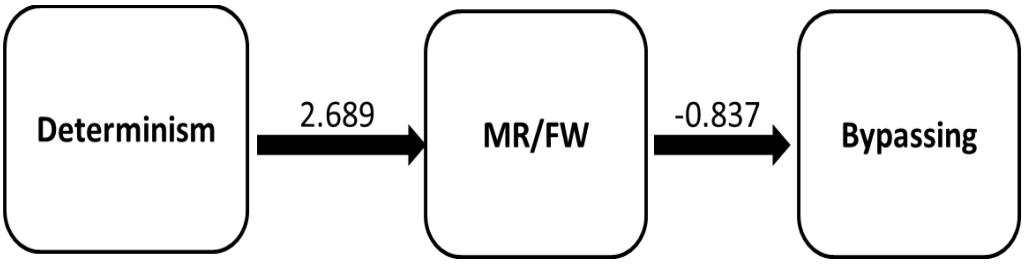


Figure 2:The MR/FW Incompatibilist Model[9]

Though the best fitting causal model of the data is the MR/FW Incompatibilist Model, it is still possible that the Bypassing Model fits the data. To test whether the Bypassing Model fits the data we constructed a structural equation model, using Tetrad IV, and found that this model did not fit the data at all.[10] This indicates that the Bypassing Model should be rejected.

4. Discussion

We tested two alternative causal models and found support for the MR/FW Incompatibilist Model: Determinism→MR/FW→Bypassing. But what we are ultimately interested in is whether the MR/FW Incompatibilist Model is the correct causal model of the data. Recall that in Section 2.3, we pointed out that Murray and Nahmias' data was indeterminate between the Bypassing and MR/FW Incompatibilist Models since the standard regression procedure used for testing mediation yielded two mediation models. We tested both the Bypassing and MR/FW Incompatibilists Models, using a procedure which provides some advantage over using a series of regressions to test causal models.

Recall that the Bypassing Model predicts that following causal model:

Determinism→Bypassing→MR/FW.

But when testing the model, we found no support for the model. The Bypassing Model ends up getting rejected. The MR/FW Incompatibilist Model, however, predicts the following causal model:

Determinism→MR/FW→Bypassing.

The data strongly support this, providing evidence that the MR/FW Incompatibilist Model is the correct causal model of our data.

[9] Each 'edge' in the model represents a direct, causal connection. The numbers above each edge are linear coefficients.

[10] χ^2 (1)=10.3052, p=.0013, BIC=6.2108.

Thus far, we've been at pains to argue that the MR/FW Incompatiblist Model provides a better explanation of bypassing results than does the Bypassing Model. As a result, we have followed Nahmias and Murray in clustering together responses to moral responsibility and free will into a single factor (MR/FW). However, to further explore what is driving the effects here, we wanted to look at these factors individually.

Perhaps the most familiar incompatibilist view is that under determinism, people are not morally responsible because they lack free will. That is, the absence of free will is what drives the absence of responsibility. In the present context, we predicted that the best causal model of the relationship between free will and moral responsibility would be this:

Determinism→FW→MR

To test this model, we began by taking responses only to the free will (FW: In Universe [A/C], it is possible for a person to have free will) and moral responsibility (MR: In Universe [A/C], it is possible for a person to be fully morally responsible for their actions) questions, treating these as our dependent variables while continuing to treat determinism as our independent variable. We then ran a causal search over the data—again using GES—and found that the following model was returned:[11]

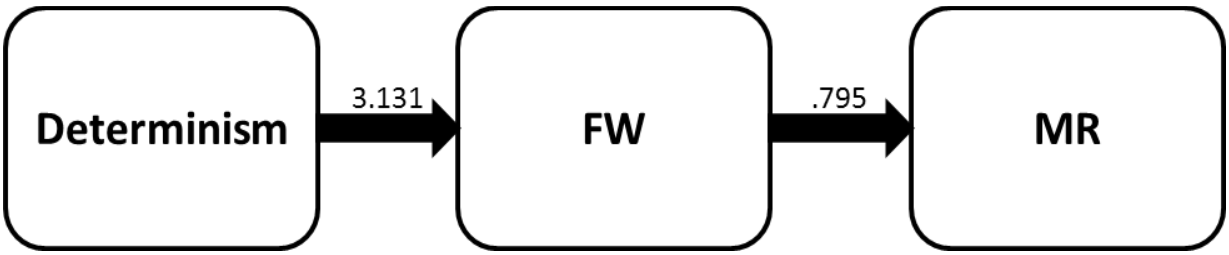


Figure 3: Relationship Between FW and MR

We also went on to construct a structural equation model, testing whether a model with MR and FW switched fit the data. But, this model ends up being rejected.[12] So, working with just MR and FW, we find support for the model whereby FW causes MR responses. This suggests that, just as incompatibilists would maintain, people think that determinism undercuts free will, which in turn undercuts moral responsibility.

We can now return to the issue of Bypassing with a somewhat more focused hypothesis. Clustering FW and MR together potentially hides important differences in how these factors relate to bypassing. So, we wanted to break apart the MR/FW cluster and test what we take to be the appropriate Incompatibilist Model with respect to Bypassing:

Incompatibilist Model: Determinism→FW→Bypassing

[11] The model returned is a good fit of the data, X^2 (1)=.8, p=.7772, BIC=-4.0143.
[12] X^2 (1)=18.2393, p<.0001, BIC=14.1450.

We ran a causal search over the data—again, using GES— looking at the relationship between Determinism, FW, and Bypassing. GES returned the Incompatibilist Model (above). This model fit the data very well.[13] And to make sure that we're not prematurely ruling out the alternative model, with FW and Bypassing switched, we constructed a structural equation model with Bypassing causing FW and tested its fit to the data. The model ends up being rejected.[14]

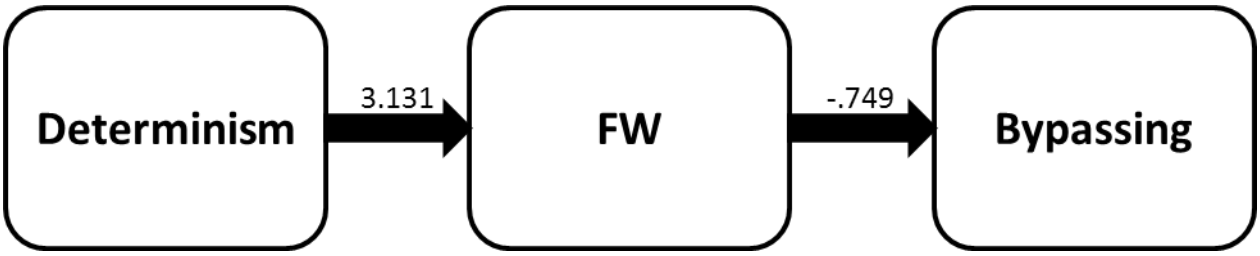


Figure 4: Relationship Between FW and Bypassing

So, we conclude that the proper model is the Incompatibilist Model.

We've seen that when using the MR/FW cluster and Bypassing cluster that the best fitting model of the data shows that MR/FW causes Bypassing (see section 3.2). And, we've seen that when breaking apart the MR/FW cluster, FW causes MR (see Figure 3) and FW causes Bypassing (see Figure 4). But, given that we've begun breaking apart the MR/FW cluster, we would like to know what the relationship is between each individual variable in the MR/FW cluster and Bypassing. So, we broke apart the MR/FW cluster, and conducted a causal search over each of the three variables (i.e., FW, MR and Blame, see Table 1) making up the MR/FW cluster and Bypassing. Using GES again, we found the following model:

[13] χ^2 (1)=.1453, p=.7031, BIC=-3.9491.
[14] χ^2 (1)=14.6579, p=.0001, BIC=10.5636.

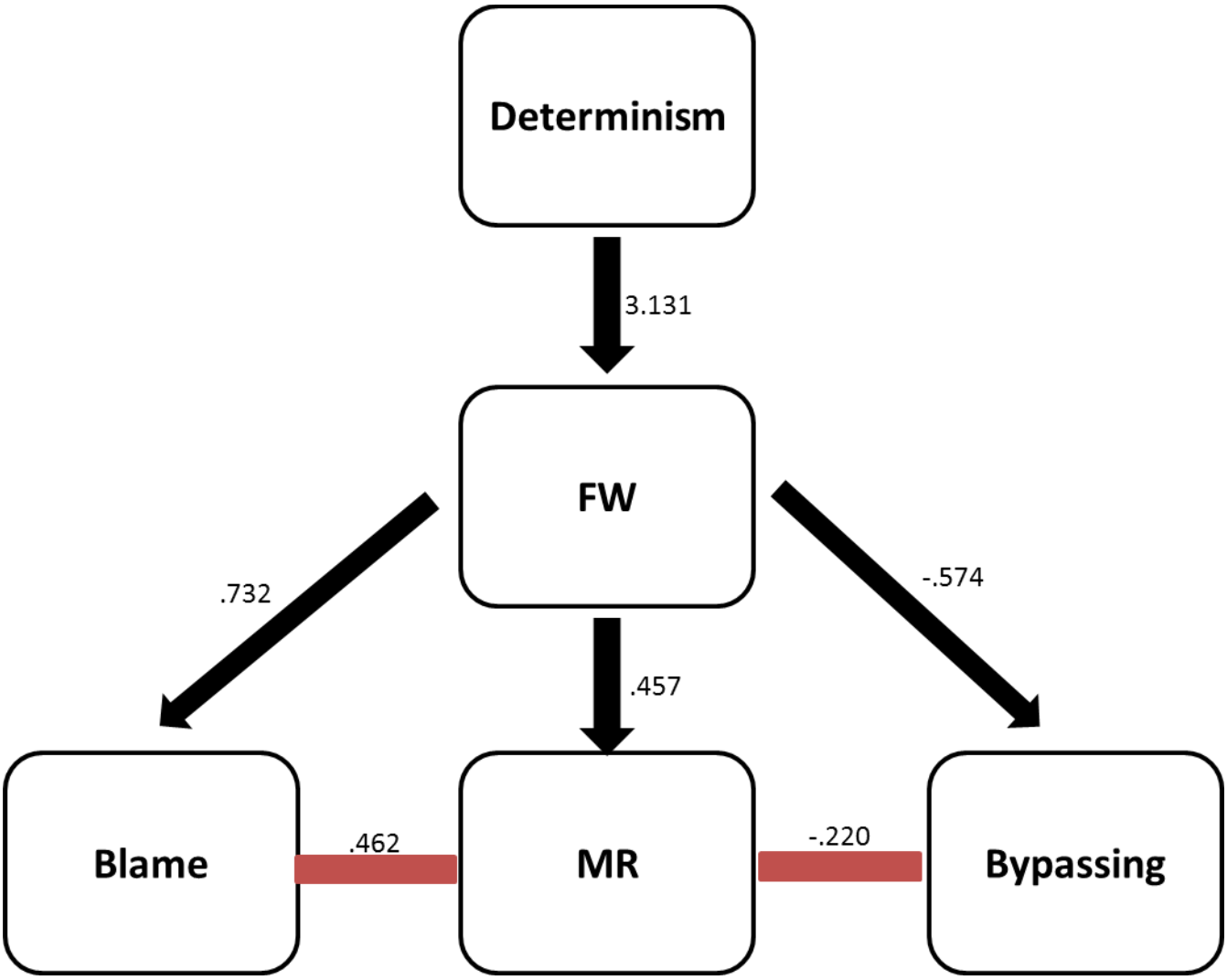


Figure 5: Overall Model Breaking Apart the MR/FW Cluster

The model returned by GES is a good fit of the data.[15] A couple of things are worth noting about the model. The first is that all of the edges were oriented by GES except for the edges between MR and Bypassing and Blame and MR. So, GES gives us no information regarding the causal relationship between MR, Bypassing, and Blame (indicated by red edges in Figure 5). Secondly, and perhaps more importantly, the above model indicates that Determinism *only* directly causes FW and it shows that FW directly causes Bypassing, Blame and MR. In short, even when we break apart the MR/FW cluster, we continue to find support for the Incompatibilist model and in particular an Incompatibilist model whereby Bypassing is largely explained by FW.

[15] X^2 (4)=.6660, p=.9555, BIC=-15.7114.

We thus take our evidence to provide strong support for the Incompatibilist Model. One might object that there could be other causal models that we failed to consider. Of course, it's true that we didn't consider other possible latent variables. Nahmias' ingenious idea was that bypassing is a mediating variable. Our goal has been to rebut this claim. We obviously can't rule out other yet-to-be-specified models. But of the models that are actually in play, we now have evidence that the Incompatibilist Model is the right causal model.

In the introduction, we noted that the bypassing account faces a puzzle. Given that descriptions of determinism don't lead people to make bypassing judgments for volcanoes or computers, why *do* they make bypassing judgments for mental states? The Incompatibilist Model, which we've argued is favored by causal modeling, provides an answer to that question. On the Incompatibilist Model, bypassing doesn't happen for volcanoes or computers because free will is not involved. So the incompatibilist view that free will doesn't exist in the determinist universe should not lead to bypassing judgments about volcanoes or computers. The bypassing judgments are caused by negative judgments concerning free will which are themselves caused by the determinist scenario. But now we face our own version of the puzzle. Why does determinism have this cascade of effects that runs through a denial of free will to an affirmation of bypassing? In particular, why would incompatibilist judgments drive bypassing judgments?

Our speculative proposal is that there is something distinctive about the way people are conceiving of *decisions* in these tasks (see also Knobe forthcoming and Sias unpublished). Participants tend to think of decisions as fundamentally indeterminist such that if determinism is true, people really don't make decisions. If that's right, the bypassing questions might provide people with a way of expressing their view that decisions don't occur under determinism.

5. Study 2: Decisions and Bypassing

For this study, we want to explore two hypotheses that flow from our proposal. Decisions are the paradigmatic form of *practical* reasoning, and our proposal is that *theoretical* reasoning will be less likely to inspire bypassing judgments. Our first prediction then, is that people will be more likely to give bypassing judgments for decisions as compared to a simple instance of theoretical reasoning – solving an arithmetic problem. Our second prediction is that people presented with a description of a determinist universe will be more likely to deny that practical reasoning occurs in that universe as compared to theoretical reasoning.

5.1. Design

We created three separate conditions:

> *Practical reasoning*: Imagine a universe 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.

Theoretical reasoning: Imagine a universe 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 was in math class and had to add 39+6 in his head. He concluded that the answer was 45. Like everything else, this reasoning was completely caused by what happened before it. So, if everything in this universe was exactly the same up until John added 39+6, then it had to happen that John would conclude that the answer was 45.

Physical event: Imagine a universe 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 an earthquake occurred. As a result, a tree fell over. Like everything else, these events were completely caused by previous events. So, if everything in this universe was exactly the same up until the tree fell, then it had to happen that the tree would fall over.

Participants were randomly assigned to one of three conditions. For each condition, participants were asked to indicate the extent to which they agreed or disagreed with one of the following statements on a 7-pt. scale anchored at 1=strongly disagree and 7=strongly agree:

Practical reasoning: "In this universe, when people make decisions, what they think and want has no effect on what actions they end up performing."

Theoretical reasoning: "In this universe, when people solve math problems the numbers they add has no effect on the answers they end up giving."

Physical event: “In this universe, the earth's shaking has no effect on whether trees fall over.”

Additionally, in each condition, participants were asked a forced choice question about the nonexistence of the relevant type of event:

Practical reasoning: In this universe, people make decisions. [Yes/No]

Theoretical reasoning: In this universe, people add numbers. [Yes/No]

Physical event: In this universe, trees fall over. [Yes/No]

Following the nonexistence question, participants were asked to indicate their degree of confidence in their answer on a 7-pt. scale anchored at 1=not at all confident and 7=extremely confident.

Finally, in each condition, participants were asked two comprehension questions:

(1) In this universe, everything is completely caused by prior events. [True/False]

(2) In this universe, everything is random. [True/False]

5.2. Results

A total of 158 participants were recruited through Amazon's Mechanical Turk. After filtering out participants who missed either one or both comprehension questions, the total number of participants was 95.

We begin with responses to the scaled questions across our three conditions. What we found was that people were more willing to express agreement with the Bypassing statement in the Practical reasoning condition (M=4.75, SD=2.17) than in the Theoretical reasoning condition (M=3.50, SD=2.34), $t(62)=2.215$, $p=.030$, or the Physical event condition (M=2.00, SD=1.86), $t(61)=5.390$, $p<.0001$. We also found that people were more willing to express agreement with the Bypassing statement in the Theoretical reasoning condition than in the Physical event condition, $t(61)=2.808$, $p=.007$, though the average response in both cases was on the side of denying bypassing. The results are shown in Figure 3.

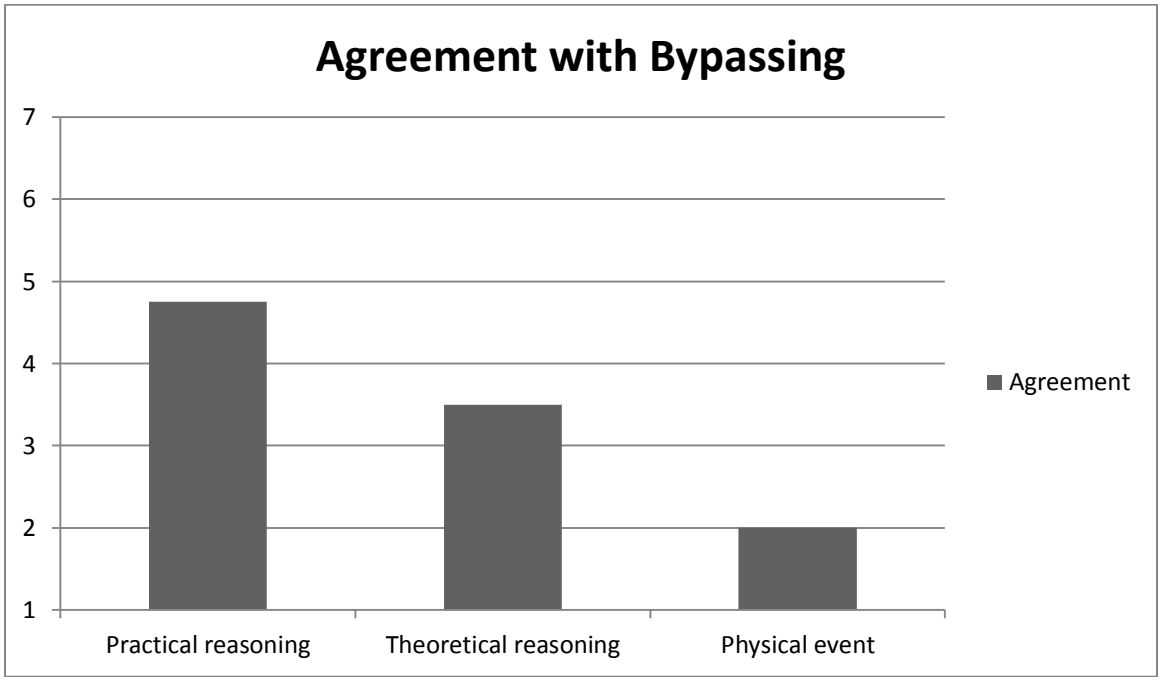


Figure 3: Mean Agreement with Bypassing Across Conditions

We then examined responses to the nonexistence questions across the three conditions and found that people were more likely to answer "no" in Practical reasoning (Yes: 53%, No: 47%) than in Theoretical reasoning (Yes: 88%, No: 12%, $\chi^2(64)=9.057$, $p=.003$) or Physical event (Yes: 100%, No:0%, $\chi^2(63)=19.072$, $p<.0001$) and that people were more likely to answer "no" in Theoretical reasoning than in Physical event ($\chi^2(63)=4.138$, $p=.042$). These results are shown in Figure 4. We also ran two regressions to examine the relationship between the nonexistence question and the bypassing question. For the theoretical reasoning case, answers on whether people add numbers in this universe significantly predicted bypassing responses $t(63)=2.17$, Beta=.369, $p=.038$. Similarly, for the practical reasoning case, answers on whether people make decisions in this universe significantly predicted bypassing responses $t(63)=4.2$, Beta=.608, $p<.001$.

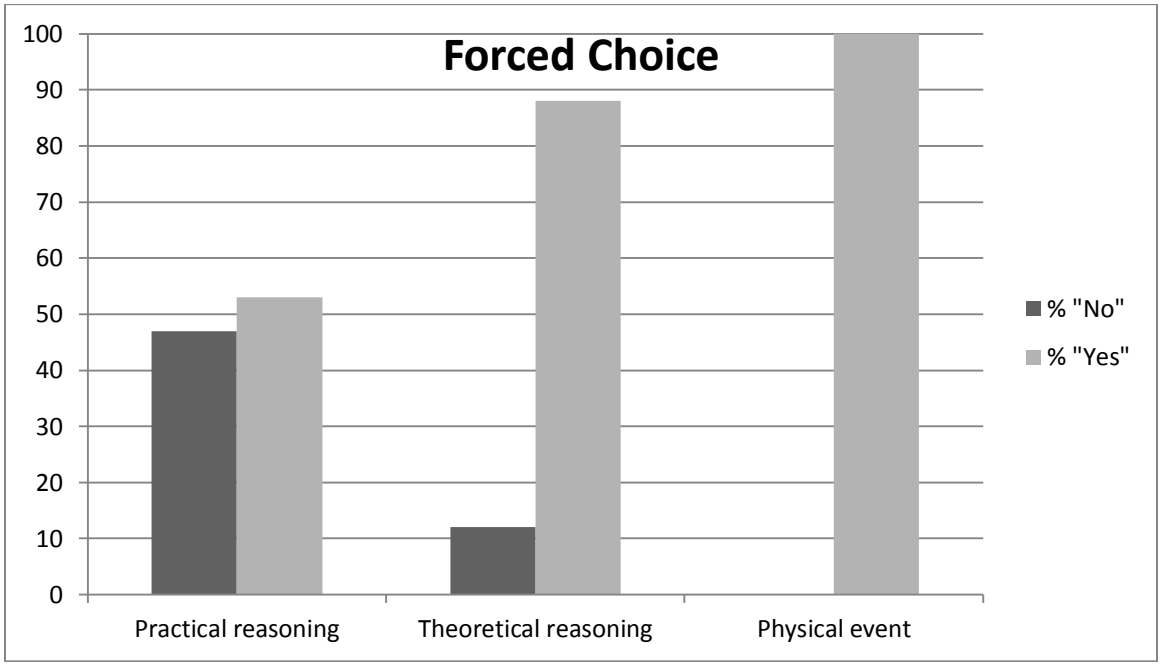


Figure 4: Forced Choice Responses to Nonexistence Question

Finally, we conducted two mediation analyses to examine the causal relationship between responses to the nonexistence and bypassing questions. Since we are ultimately interested in whether a denial of human decision making is causally responsible for bypassing judgments we restricted our analyses to responses in both the Practical reasoning and Theoretical reasoning cases. Thus our Reasoning variable included only the Practical reasoning and Theoretical reasoning cases and our Nonexistence and Bypassing variables included only responses in the Practical reasoning and Theoretical reasoning cases.

The first mediation analysis we ran was aimed at testing the following causal model:

Reasoning→Nonexistence→Bypassing.

We found that Nonexistence fully mediates responses to Bypassing. That is, once we factor in how people respond on the Nonexistence question, the type of Reasoning does not predict Bypassing responses.[16] We then ran a mediation analysis testing the other possible model:

Reasoning→Bypassing→Nonexistence.

Here we found that responses to Bypassing did *not* mediate responses on Nonexistence.[17] To corroborate these findings, we also constructed two structural equation models. As in study 1, we used GES[18] to search over our data and found that the best fitting causal model of the data was the Reasoning→Nonexistence→Bypassing.[19] We then tested the alternative causal model—Reasoning→Bypassing→Nonexistence—finding that it does not fit the data at all and thus should be rejected.[20]

---

[16] We began by following the procedure outlined in Baron and Kenny (1986) for testing mediation:

Step 1: A regression model with Reasoning as a predictor of Bypassing was significant, t(63)=-2.215, Beta=-.271, p=.030.

Step 2: A logistic regression model with Reasoning as a predictor of Nonexistence was significant, Wald X^2 (63)=8.062, B=-1.821, p=.005.

Step 3: A regression model with Nonexistence as a predictor of Bypassing was significant, t(63)=5.078, Beta=.543, p=.000.

Step 4: When Reasoning and Nonexistence were both included in a multiple regression model, the effect of Reasoning on Bypassing was not significant, t(63)=-.671, Beta=-.078, p=.505.

Note that the Baron and Kenny steps are only the beginning. When using only continuous data and a binary predictor variable X, one can simply test mediation using the standard Baron and Kenny procedure along with a Sobel (1982) test. But if either the mediator variable M or outcome variable Y are *dichotomous*, one ends up using logistic and linear regression analyses and so the coefficients for the dichotomous and continuous variables are in different scales and thus we can't simply "mash them into a z-test" (Iacoubucci, p.8, 2012). To handle this, the coefficients and standard errors need to be rescaled. We will defer a detailed discussion of this procedure (though the interested reader should see Herr, 2013 for details; also see Iacoubucci, 2012) and simply note that we followed the procedure outlined in Herr (2013). We found that the reduction in the effects of Reasoning on Bypassing when Nonexistence was included in the model was significant *Z*=-2.3932, p=.016. This suggests that Nonexistence *fully* mediates the effects of Reasoning on Bypassing.

[17] Again, we began by following the procedure outlined in Baron and Kenny (1986) for testing mediation:

Step 1: A logistic regression model with Reasoning as a predictor of Nonexistence was significant, Wald X^2 (63)=8.062, B=-1.821, p=.005.

Step 2: A regression model with Reasoning as a predictor of Bypassing was significant, t(63)=-2.215, Beta=-.271, p=.030.

Step 3: A logistic regression model with Bypassing as a predictor of Nonexistence was significant, Wald X^2(63)=12.625, B=.742, p=.000.

Step 4: When Reasoning and Bypassing were both included in a multiple logistic regression model, the effect of Reasoning on Nonexistence was still significant, Wald X^2 (63)=5.071, B=-1.688, p=.024.

Since the effect of Reasoning remained significant even when Bypassing was included in the model, step 4 was not satisfied. Nonetheless, we followed the procedure in Herr (2013) for rescaling, finding that the reduction in the effects of Reasoning on Nonexistence when Bypassing was included in the model was not significant Z=-1.8322, p=.067. This suggests that Bypassing does *not* mediate the effects of Reasoning on Nonexistence.

[18] As is standard, binary variables can be interpreted as continuous when using GES. But, we note that this only works if the binary variables are given values of 0 and 1 when interpreted as continuous. Thanks to Joe Ramsey for guidance on this.

[19] This model was a very good fit of the data, χ^2 (1)=.4635, *p*=.4960, BIC=-3.6954.

[20] χ^2 (1)=5.280, *p*=.0216, BIC=1.1211.

6. Discussion

As expected from previous work, participants make bypassing judgments about decisions but not about physical processes like trees falling. More importantly, we found that people tended not to make bypassing judgments about the case of theoretical reasoning we used. When it comes to simple mental arithmetic, people tend to *deny* that bypassing occurs. Turning to the nonexistence question, nearly half of the subjects denied that decisions happen in the determinist universe.[21] In addition, people's responses to the question about whether decision exists in a determinist universe was an excellent predictor of how they would respond to the bypassing question. People who denied the existence of decisions under determinism gave higher bypassing scores. Moreover, we found unique support for the Reasoning→Nonexistence→Bypassing causal model. This provides strong support for our speculative proposal: the reason a deterministic scenario has the cascade of effects culminating in bypassing responses is that many people think that if determinism is true, people don't even make decisions.

This result also helps to explain another finding in Murray & Nahmias' paper. In their paper, Murray and Nahmias attempted to intervene directly on bypassing by presenting participants with a description of determinism coupled with a denial of bypassing. In this set up, people gave higher FW/MR judgments. At first glance, this suggests that bypassing really is driving incompatibilist judgments. However, a closer look at the results again reveals something that, by our lights, is more interesting. When participants were presented with the Nichols & Knobe description of determinism that we've been using in our studies, even adding the Bypassing clarification only reduced Bypassing judgments from 66% to 49% in their studies (see Murray and Nahmias, forthcoming, Table 1). Half of the subjects *continued* to say that bypassing occurred. As Murray & Nahmias note, this suggests that this description of determinism easily invites bypassing responses: "We believe these results further suggest that N&K's description of determinism is easily read to involve bypassing, such that when participants are also *told* that it does not involve bypassing, they may simply be unsure about how to interpret the seemingly conflicting information that they are given about the scenario" (Murray & Nahmias, MS, fn 27). A scenario in which causal determinism - but *not* bypassing - is supposed to hold presents a package that is intuitively difficult for participants to process. Our studies provide an explanation for this. It's not that the description of determinism directly leads people to bypassing judgments. Rather, the description of determinism leads people to think that decisions don't occur and this denial of decisions is what leads to the affirmation of bypassing.

7. Study 3: Theoretical and Practical Reasoning

In study 2, we attempted to show a difference in how people thought about theoretical versus practical reasoning under determinism. We wanted to show that the prospect of determinism does not lead people to deny the existence of simple kinds of theoretical reasoning, like simple mental arithmetic. However, there was an important asymmetry between our case of theoretical reasoning and our case of practical reasoning. In the practical reasoning case, we simply said

[21] An important precedent for this result comes from unpublished work by Jim Sias. He described a deterministic universe to participants and asked whether it made sense to say that the agent in the scenario was able to make decisions. Sias found that participants tended to deny that it made sense (Sias, unpublished).

that John made a decision to have French fries. By contrast, for the addition case, we provided a more detailed process story – we said that John added 39 + 6 and concluded with the answer 45.[22]

To address this potential confound, we ran a new study focused just on whether people would be more likely to allow that a determined agent engages in simple theoretical reasoning than practical reasoning. We designed the cases to be much closer matches. First, we provided more context for the decision by noting the available options. Second, we described the situation that the agent faced in each case and the agent's response, but we did not explicitly say anything about the process the individual underwent.

7.1. Design

We created two cases:

> *Practical Reasoning*: Imagine a universe 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 was in a restaurant, looking at the menu. There were several items on the menu, including cole slaw, French fries, and fruit. When the server asks what he would like, John said "French fries". Like everything else, John's saying "French fries" was completely caused by what happened before it. So, if everything in this universe was exactly the same up until John said "French fries", then it had to happen that John would say "French Fries".
>
> *Theoretical Reasoning*: Imagine a universe 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 Mark was in a restaurant, looking at the menu. He sees that the French fries cost $4, fruit costs $5, and a soda costs $2. A friend asks Mark how much French fries and a soda costs. Mark says, "$6." Like everything else, Mark's saying "$6" was completely caused by what happened before it. So, if everything in this universe was exactly the same up until Mark said "$6" then it had to happen that Mark would say "$6".

After reading each case, participants were asked to explain, in their own words, what they had just read. Following each story, participants were asked to indicate their agreement, on a 6-point scale (anchored at 1=strongly disagree and 6=strongly agree), with the following statements regarding whether the agent engaged in the relevant kind of reasoning:

> *Practical Reasoning Probe*: Before John said "French fries", he made a decision.
>
> *Theoretical Reasoning Probe*: Before Mark said "$6", he added the prices.

For each case, after answering this question, participants had to answer a control question:

[22] We thank a referee for pointing out the asymmetry and for prompting us to run this additional task.

Practical Reasoning Control: According to the story, it had to happen that Mark would say "French fries". [yes/no]

Theoretical Reasoning Control: According to the story, it had to happen that Mark would say "$6". [yes/no]

We used a within-subjects design, counterbalanced for order, with participants being randomly assigned to one of the two orders.

7.2. Results

A total of 107 participants were recruited through Amazon's Mechanical Turk. Two people were removed from the data for missing a control question, and two more people were removed for failing to provide an explanation of what they read (e.g., "I don't know" and "I didn't read it"). The remaining number of participants was 103.

Since in each order, participants were presented with either the Practical or Theoretical Reasoning case first, we first analyzed the difference between those first responses with a between subjects *t*-test. There was a significant difference ($t(101)=-2.429$, $p=.017$) between those that read the Practical Reasoning ($M=3.94$, $SD=1.83$) and those that read the Theoretical Reasoning ($M=4.77$, $SD=1.65$) cases first on whether the agent engaged in the relevant kind of reasoning. The results are shown in Figure 5.

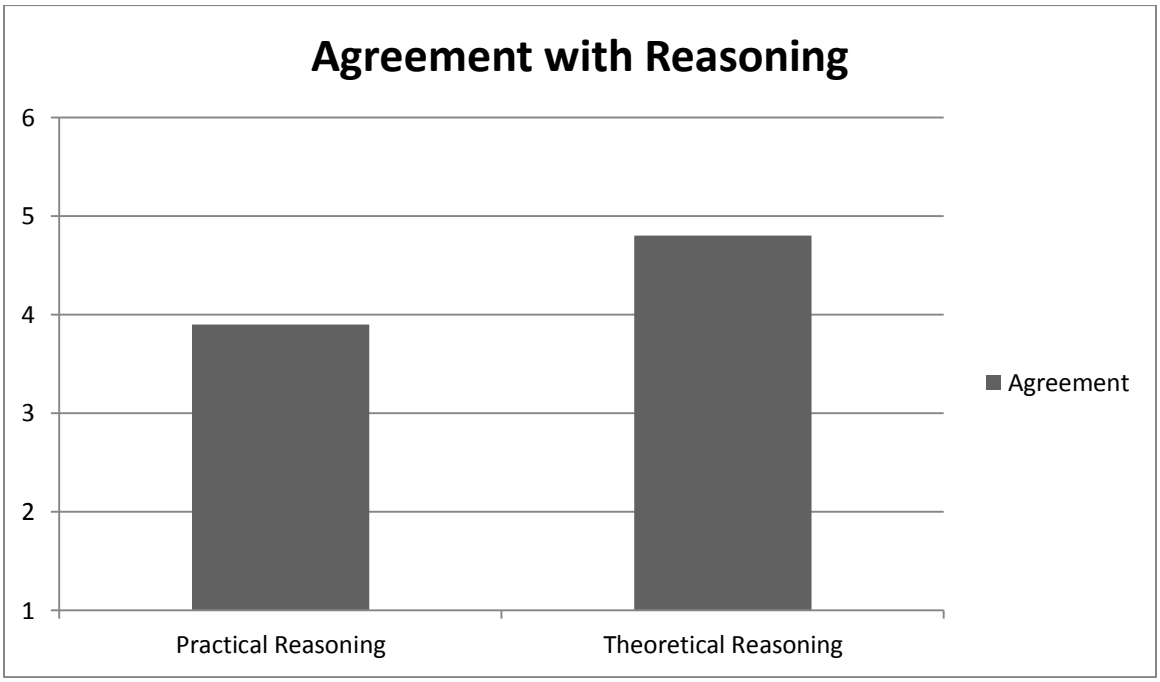


Figure 5: Agreement with Reasoning Question

We also conducted a within-subjects test. But before testing for a within-subjects effect, we checked for order effects. There were no effects of order on either the Practical (Presented First, M=3.94, SD=1.83; Presented Second, M=4.13, SD=1.86, t(101)=-.527, p=.599) or Theoretical (Presented First, M=4.66, SD=1.72; Presented Second, M=4.78, SD=1.66, t(101)=-.384, p=.702) Reasoning Probes. Since there were no order effects, we combined the data from both orders before conducting the within-subjects analysis. The within-subjects test revealed a significant difference between Practical (*M*=4.07, *SD*=1.83) and Theoretical Reasoning (*M*=4.69, *SD*=1.67) on the reasoning question (*t*(105)=3.349, *p*=.001).

With the cases more closely matched, we continue to find evidence of the same asymmetry that was found in study two: under the stipulation of determinism, people are more likely to withhold attribution of practical reasoning than simple instances of theoretical reasoning. Thus, the results of both studies 2 and 3 suggest that, under determinism, people tend to allow for the presence of simple instances of theoretical reasoning while they are more likely to withhold attributions of decision making. We take this as further support for our speculative proposal that many people think that if determinism is true, this challenges the very idea that people make decisions. And this is what explains the apparent effect of bypassing responses in deterministic scenarios.

8. The depth of incompatibilism

As we noted at the beginning of the paper, intuitions play a central role in motivating incompatibilism. In the face of the Bypassing challenge, we have argued that people do indeed have intuitions that cohere with incompatibilism. But the results of investigating bypassing intuitions reveals something of further significance. The incompatibilist claim is not simply that free choices aren't determined. For one might well hold that free choices happen to be undetermined, but this isn't a critical feature of free choice. In order to hold that free will is actually inconsistent with determinism, the incompatibilist needs to defend the idea that indeterminism is in fact a critical feature of free choice. Our studies provide some evidence that people regard free will as critically entwined with indeterminism.

Several studies (Nahmias et al. 2007; Nahmias & Murray 2010; Murray & Nahmias forthcoming) have shown that when presented with determinism, people give bypassing judgments. For instance, when given a description of a deterministic universe, participants claim that in that universe what people think and want have no effect on what they end up doing. Although this seems like an outright confusion, the pattern of results reveal something novel about ordinary views of choice. First, we found that the best causal model for the phenomenon is not (*pace* Murray and Nahmias) that determinism leads people to infer bypassing which then leads people to deny free will. Rather, the best model is that determinism leads people to deny free will and this then leads to the affirmation of bypassing. Furthermore, a key part of the explanation for the affirmation of bypassing is that many people think that determinism precludes decision making. It's not that they think decision occurs in some way that bypasses beliefs and values; they think that decision doesn't occur at all. Now, as philosophers we are so familiar with compatibilist models of decision making (e.g., Nelkin 2004, Pereboom 2008) that we have difficulty recapturing the pre-theoretic innocence that lay people bring to these questions. But perhaps we can get a feel for it as follows. Assume that determinism is true and consider whether you agree with the following: "Random processes have no effect on people's

behavior." This statement seems right, under the assumption of determinism. For if determinism is true then random processes don't affect behavior because there are no random processes.

As philosophers we can easily see the theoretical usefulness of the notion of decision even under the assumption of determinism. For instance, we have natural ways of construing decision making in terms of calculating over subjective utility. It is an open empirical question how easily ordinary people can be led to acknowledge a notion of decision making that is not beholden to indeterminism. In any case, however, the bypassing experiments do not provide a context in which ordinary people think about decisions in a way that is neutral about determinism. Rather, the bypassing experiments seem to elicit a way of thinking about decision that is antithetical to determinism. Thus, instead of showing that folk incompatibilism is an error, the bypassing results help to show just how deep incompatibilism goes. As noted above, according to incompatibilism, the idea that free choices aren't determined is not just a peripheral fact about free choice – it is at the core of the idea of free choice. The results on bypassing suggests that determinism does indeed strike to the core of the way people ordinarily think about free choice. Indeed, it strikes to the core of the way many people think about choice, *simpliciter*. That, we suggest, is the real lesson of bypassing.

References

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182

Chan, H. forthcoming. Incompatibilist about what?

Chickering, D. (2002). Optimal structure identification with greedy search. *Journal of Machine Learning Research* , 3, 507 - 554.

Deery, O., Bedke, M. and Nichols, S. forthcoming. "Phenomenal Abilities: Incompatibilism and the Experience of Agency." In D. Shoemaker (ed.), *Oxford Studies in Agency and Responsibility*.

Herr, N. (2013). Mediation with Dichotomous Outcomes. http://www.nrhpsych.com/mediation/logmed.html

Holton, R. (2009). "Determinism, Self-Efficacy, and the Phenomenology of Free Will." Inquiry: An Interdisciplinary Journal of Philosophy 52.4: 412-428.

Iacobucci, D. (2012). Mediation analysis and categorical variables: The final frontier. *Journal of Consumer Psychology*. doi:10.1016/j.jcps.2012.03.006

Iacobucci, D., Saldanha, N., and Deng, X. (2007). A Mediation on Mediation: Evidence That Structural Equation Models Perform Better Than Regressions. *Journal of Consumer Psychology*, 17(2), 140-154.

Kane, R. (1999). Responsibility, luck, and chance: reflections on free will and indeterminism. *Journal of Philosophy*, 96: 217-240.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), The handbook of social psychology (Vol. 1, 4th ed., pp. 233–265). Boston: McGraw-Hill.

Kenny, D. (2012). Mediation. http://davidakenny.net/cm/mediate.htm

Knobe, J. (forthcoming). Free Will and the Scientific Vision.

Meek, C. (1997). Graphical Models: Selecting causal and statistical models. PhD Thesis, Carnegie Mellon University.

Mele, A. (forthcoming). Free will and substance dualism: The real scientific threat to free will?

Murray, D. & Nahmias, E. (forthcoming). Explaining Away Incompatibilist Intuitions. *Philosophy and Phenomenological Research*.

Nahmias, E., Coates, D. J., & Kvaran, T. (2007). Free will, moral Responsibility, and mechanism: Experiments on folk intuitions. *Midwest Studies in Philosophy*, 31: 214-242.

Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2006). Is incompatibilism intuitive? *Philosophy and Phenomenological Research*, 73: 28-53.

Nahmias, E., & Murray, D. (2010). Experimental philosophy on free will: An error theory for incompatibilist intuitions. In *New Waves in Philosophy of Action*, ed. by J. Aguilar, A. Buckareff, and K. Frankish. Palgrave-Macmillan, 189-215.

Nelkin, Dana. 2004. "Deliberative alternatives," *Philosophical Topics* 32: 215–240.

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

Pereboom, D. (2001). *Living without free will*. Cambridge: Cambridge University Press.

Pereboom, Derk. 2008. "A Compatibilist Account of the Epistemic Conditions on Rational Deliberation," *The Journal of Ethics* 12/3: 287–306.

Rose, D., Livengood, J., Sytsma, J., & Machery, E. (2011). Deep trouble for the deep self. *Philosophical Psychology*, 25(5), 629-646.

Roskies, A. & Nichols, S. (2008). Bringing moral responsibility down to earth. *Journal of Philosophy*, 105(7): 371-388.

Sarkissian, H., Chatterjee, A., De Brigard, F., Knobe, J., Nichols, S., and Sirker, S. 2010. "Is Belief in Free Will a Cultural Universal?" *Mind & Language*, 35, 346–358.

Sias, J. unpublished. Decisions, decisions: A study of folk intuitions.

Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In S. Leinhart (Ed.), Sociological methodology (pp. 290–312). San Francisco: Jossey-Bass.

Sommers, T. (2012). *Relative Justice.* Princeton: Princeton University Press.

Strawson, G. (1986). *Freedom and belief*. Oxford: Clarendon Press.

Van Inwagen, P. (1983). *An Essay on Free Will.* Oxford University Press.