Freedom from what? Separating lay concepts of freedom

Claire Simmons [a,b,*], Paul Rehren [c], John-Dylan Haynes [d,e], Walter Sinnott-Armstrong [a,b,f]

[a] *Duke University Kenan School of Ethics, 1364 Campus Dr, Durham, NC 27705, USA*
[b] *Duke Institute for Brain Sciences, 308 Research Dr, Durham, NC 27710, USA*
[c] *Ethics Institute, Department of Philosophy and Religious Studies, Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, the Netherlands*
[d] *Max Plank Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany*
[e] *Charité – Universitätsmedizin Berlin, Bernstein Center for Computational Neuroscience, Haus 6, Philippstrasse 13, 10115 Berlin, Germany*
[f] *Duke University Department of Philosophy, 1364 Campus Dr, Durham, NC 27705, USA*

A B S T R A C T

Debates about freedom of will and action and their connections with moral responsibility have raged for centuries, but the opposing sides might disagree because they use different concepts of freedom. Based on previous work, we hypothesized that people who assert freedom in a determined (D) or counterfactual-intervener (CI) scenario assert this because they are thinking about freedom from constraint and not about freedom from determination (in D) or from inevitability (in CI). We also hypothesized that people who deny that freedom in D or in CI deny this because they are thinking about freedom from determination or from inevitability, respectively, and not about freedom from constraint. To test our hypotheses, we conducted two main online studies. Study I supported our hypotheses that people who deny freedom in D and CI are thinking about freedom from determinism and from inevitability, respectively, but these participants seemed to think about freedom from constraint when they were later considering modified scenarios where acts were not determined or inevitable. Study II investigated a contrary bypassing hypothesis that those who deny freedom in D denied this because they took determinism to exclude mental causation and hence to exclude freedom from constraint. We found that participants who took determinism to exclude freedom generally did not deny causation by mental states, here represented by desires and decisions. Their responses regarding causation by desires and decisions at most weakly mediated the relation between determinism and freedom or responsibility among this subgroup of our participants. These results speak against the bypassing hypothesis and in favor of our hypothesis that these participants were not thinking about freedom from constraint.

1. Introduction

In 1924, Nathan Leopold and Richard Loeb kidnapped and murdered 14-year-old Bobby Franks. They confessed and pled guilty, but the famous lawyer Clarence Darrow saved them from the death penalty by arguing (for 12 h!):

> I know they cannot feel what you feel and what I feel, that they cannot feel the moral shocks which come to men who are educated and who have not been deprived of an emotional system or emotional feelings…. Is Dickie Loeb to blame because out of the infinite forces that conspired to form him, the infinite forces that were at work producing him ages before he was born, that because out of these infinite combinations he was born without it? If he is, then there should be a new definition for justice. (Darrow 1924, 220-221)

---

* Corresponding author at: Kenan Institute for Ethics Duke University Durham, Box 90432, NC 27705, USA.
 E-mail address: cs568@duke.edu (C. Simmons).

> I am trying to say to this court that these boys are not responsible for this, and that their act was due to this and this, and this and this, and asking this court not to visit the judgment of its wrath upon them for things for which they are not to blame. (Darrow 1924, 354)

Darrow's argument seems to assume that agents are not responsible or to blame for their acts if they "cannot" feel or do otherwise or if those acts are determined or caused by "forces" "before [they were] born".

Something like that assumption is shared by many philosophers and scientists, who are called incompatibilists because they hold that determinism is incompatible with freedom and responsibility. (For an overview, see Clarke & Capes 2020.) But why exactly is determinism supposed to exclude freedom? Some incompatibilists hold that determinism precludes freedom because it makes the agent unable to do otherwise or makes the act inevitable (e.g. van Inwagen, 1983). Others argue that determinism precludes freedom because it implies that all acts are caused, and a caused act cannot be free (e.g. Ginet 1997; Pereboom 2007; McCann 2012).

Another group of philosophers disagrees and claims that determinism can co-exist with freedom (e.g. Dennett 1984). On this so-called compatibilist view, some causes (like physical restraint, mental illness, and coercion) do remove freedom, but other causes (such as normal desires, incentives, or neural activity) do not remove or even reduce freedom, presumably because those kinds of causes do not prevent agents either from doing what they really want to do or from responding to reasons. Because determinism does not imply that all acts are caused by the limited kinds of causes that remove freedom, compatibilists think that freedom and determination can co-exist. (For an overview, see McKenna & Justin Coates, 2020.).

The compatibilist and incompatibilist camps in philosophy seem to be talking about freedom *from* different kinds of barriers: determination, inevitability, causation, and constraint (Sinnott-Armstrong, 2012). Insofar as a particular concept of freedom is defined by which barriers remove that kind of freedom, these philosophers use different concepts:

- *Freedom from determination*: when the act is not determined
- *Freedom from causation*: when the act is not caused
- *Freedom from inevitability*: when the act is not inevitable (the agent can do otherwise)
- *Freedom from constraint*: when the agent is not constrained or prevented from doing what they really want to do (or prevented from responding to reasons)

For example, suppose a thief threatens to shoot an agent if they do not hand over their car keys. The agent is not free from *causation* (because the threat causes the action) or from *constraint* (because coercion is a constraint) but is still free from *inevitability* (if the agent is able to run away without being shot, despite the risks). On the other hand, if the car owner hands the keys to a stranger because they want to collect insurance and buy a new car, then they are not free from *causation* (because their desire causes their action), but they are free from *constraint* (because their desires do not constrain them from doing what they want) and also free from *inevitability* (because they are able to keep their keys to themselves).

Which concept is common among lay persons? Of course, people likely hold different beliefs about freedom for different reasons, but surveys can be used to find trends. Some psychologists suggest that most people think of free action and will only in terms of freedom from constraint, not freedom from causation, inevitability, or determination (in line with compatibilists like Dennett). Several studies support this view. One study asked participants to define free will and concluded, "the folk concept of free will is defined by the capacity to choose based on one's desires and free from constraints" (Monroe & Malle 2010, 25). Another study reached a similar conclusion by asking participants to write autobiographical narratives of free and unfree actions (Stillman, Baumeister, & Mele 2011). Again, participants in Vonasch, Baumeister, & Mele (2018) tended to say that agents are free unless their ability to make decisions is limited by certain internal factors (such as mental illness or addiction) or external factors (such as social pressure or being drugged). And in Monroe, Dillon, & Malle (2014), only the extent to which different kinds of agents (e.g. human, cyborg, advanced robot) were seen as having the ability to make unconstrained choices predicted whether the agent was seen as free.

However, a different picture emerges when people judge acts in concrete scenarios. In a large cross-cultural study, Hannikainen et al. (2019) asked participants to read two scenarios. In the first scenario (D for determined), an agent in an explicitly deterministic universe decides to murder his family because he wants to run away with his secretary. In the second scenario (CI for counterfactual intervener), scientists implant a chip in the agent's brain that takes away the agent's ability to do otherwise (specifically, his ability not to kill his friend), because it would stop the agent from doing otherwise if he tried (that is, it would make him kill his friend if he decided not to). This agent decides on his own to kill his friend, so the scientists never actually interfere.

The two scenarios differ in several respects, which makes comparisons difficult, but they exemplify loss of freedom from determination (D) and freedom from inevitability (CI) but not freedom from constraint. People who think of freedom of action as freedom from constraint would be expected to judge that the agents in the D and CI scenarios acted freely, because the agents in both scenarios act on their desires and are not constrained (at least insofar as the scientists do not actively intervene in CI). However, contrary to the research cited above, only 47% of participants asked about the D scenario answered that the agent acted freely (see also Feltz, Cokely, & Nadelhoffer 2009), so around half said that the agent did *not* act freely. Regarding the CI scenario, 82% answered that the agent acted freely, so 18% said that the agent did *not* act freely. For all of those who denied that either agent acted freely, freedom from constraint is *not* enough for free action. They require something else.

What is this something else? In the D scenario, the simplest answer is freedom from *determination*. After all, these participants denied freedom in a universe where everything (including the agent murdering his family) is determined. In the CI scenario, the simplest answer is that participants are thinking about the ability to do otherwise. In other words, when they deny freedom in the CI scenario, they do so because they think that free action requires freedom from *inevitability*.

However, these simple answers might not paint the full picture. People might think that determinism rules out freedom *because*

determinism implies universal causation, and free action requires freedom from *causation* (cf. Ginet 1997; McCann 2012). A more complex possibility is that people (mistakenly) take determinism to guarantee that every event would have happened no matter what the agent decided, planned, wanted, believed, or tried to do, so our mental states (i.e. our decisions, desires, beliefs, plans, etc.) never cause our actions (Murray & Nahmias 2014; Nahmias & Murray 2011). Instead, on this view, the deterministic causal chains that produce our decisions and actions contain only physical causes and *bypass* our mental states (cf. the Bypassing Hypothesis discussed below). If so, and if these people also assume that agents are constrained unless they can do what they want *because* they want to do it, an unconstrained act must be caused by a mental state, then they might think that determinism implies that every act is constrained in this way. A similar story can be told about the CI scenario. If they follow this line of reasoning, then people who deny freedom in the D and CI scenarios might be thinking of freedom of action as freedom from *constraint*.

To summarize, we distinguished four concepts of freedom: freedom from determination, freedom from inevitability, freedom from causation, and freedom from constraint. We argued that it is not yet clear which concept or concepts of freedom are used by people who deny that any determined act is free or whether they use the same concept as others who affirm that some determined acts are free. Similarly, it is not clear whether people who deny that an agent who cannot do otherwise acts freely use the same concept of freedom as others who allow that an agent sometimes acts freely without being able to do otherwise. These apparent opponents might disagree only because they are using different concepts of freedom (see Chalmers, 2011 for a rich account of conceptual disputes).

Finding out which concept of freedom different people use in different cases might enable us to make some progress in ancient philosophical disputes about free action and free will and to test psychological claims about the lay understanding(s) of freedom (such as Monroe & Malle, 2010; Monroe, Dillon, & Malle, 2014; Stillman, Baumeister, & Mele, 2011; and Vonasch, Baumeister, & Mele, 2018). That is our goal in this study. When some people say that agents in the D or the CI scenario act freely, is this because they think freedom of action requires only freedom from constraint? And when other people say that agents in the D or CI scenarios do *not* act freely, is this because they are thinking instead about freedom from determination, freedom from causation, freedom from inevitability, or freedom from constraint?

In order to answer these questions, we need some way to determine which kind of freedom participants are thinking about in various cases. Our proposal is that people's concepts of freedom can be revealed by their responses to questions about which changes in the situation affect whether or not the agent acts freely. People who are thinking about freedom from *determination* should deny that an agent ever acts freely in a deterministic universe but assert that an agent sometimes acts freely in an alternative universe that is not deterministic. People who are thinking about freedom from *causation* should deny that an agent ever acts freely when the act is caused by anything. People who are thinking about freedom from *inevitability* should deny that an agent ever acts freely when the agent cannot do otherwise but sometimes acts freely when the barrier to doing otherwise is removed. And people who are thinking about freedom from *constraint* should assert that an agent acts freely even when the agent's act is caused or inevitable if the cause of the action is not a constraint that prevent the agents from responding to reasons and doing what they really want to do.

Based on previous work, we hypothesize that people who assert that agents in the D or the CI scenario act freely say this because they are thinking about freedom from constraint and not freedom from determination (in D) or inevitability (in CI). We also hypothesize that people who deny that the agent freely acts in D or in CI deny this because they are thinking of freedom from determination or freedom from inevitability, respectively, and not freedom from constraint. We make no prediction about how many people think about freedom from causation in these scenarios. Thus, we have four hypotheses with two parts each:

> (H1) People who *affirm* that an agent acts freely in a scenario where the act is determined (D) are mostly thinking (a) about freedom from constraint and (b) not about freedom from determination.
> (H2) People who *deny* that an agent acts freely in a scenario where the act is determined (D) are mostly thinking (a) about freedom from determination but (b) not about freedom from constraint.
> (H3) People who *affirm* that an agent acts freely in a scenario where the agent cannot do otherwise (the CI scenario) are mostly thinking (a) about freedom from constraint and (b) not about freedom from inevitability.
> (H4) People who *deny* that an agent acts freely in a scenario where the agent cannot do otherwise are mostly thinking (a) about freedom from inevitability and (b) not about freedom from constraint.

Our study has two subsidiary goals. Although our main concern is freedom of action, we also want to investigate freedom of will. Philosophers sometimes distinguish these two concepts (Gert & Duggan, 1979). It is not clear whether the same concept of freedom applies to both actions and wills or choices. A barrier (perhaps like the scientists' brain chip in CI) that makes an agent unable to act or do otherwise might not also make the agent unable to choose, decide, or will otherwise. Thus, we want to determine whether lay persons draw this distinction and whether they use the same concept of freedom for acts and for wills.

In addition, we want to explore whether people take freedom to be required for moral responsibility. Freedom is important at least partly because of its relation to responsibility, and many people seem to assume that an agent cannot be responsible for an act (or its consequences) unless the agent did that act freely (Wolf, 1990, 3-4). However, some recent studies have cast some doubt on this supposedly obvious entailment (Figdor & Phelan, 2015; Vierkant et al., 2019). What is still not clear, however, is which kind of freedom people are thinking about when they assert or deny that responsibility requires freedom.

To test our hypotheses, we conducted a pilot and two main online studies. The pilot study was used to inform the questions asked in the main studies. Here we will report the main studies as Studies I and II.

2. Study I

2.1. Materials and methods

2.1.1. Participants

Participants were recruited through the online subject pool Prolific (https://www.prolific.co.uk). We restricted participation to US residents whose first language was English, had completed at least 50 other studies in Prolific, and had at least a 90% acceptance rating on Prolific. Participants were compensated above minimum wage for their time, and our study was approved by the campus IRB of Duke University.

We recruited a sample of 400 participants and then excluded all participants who failed to respond accurately to at least one comprehension check, described in Measures below. (None of our reported results change significantly if these participants are included.) Our final sample included $n = 342$ participants (55.9% Female, 41.9% Male, 1.3% Other, 0.8% Prefer not to answer; 23.9% 18–24 years, 36.6% 25–34 years, 18.0% 35–44 years, 11.3% 45–54 years, 7.0% 55–64 years, 3.2% 65 years and over).

Participants were contacted by Prolific with an invitation to participate in the study, which was hosted on Qualtrics (https://www.qualtrics.com/). Participants gave informed consent, received instructions, and completed the study materials. Each participant completed both the D and the CI scenario, in random order. Participants then provided demographic information and exited the survey.

2.1.2. Design

Because our hypotheses (H1)-(H4) claim that people who affirm that an agent acts freely use a different concept of freedom than those who deny that an agent acts freely, we needed a branching design to separate those who affirm freedom from those who deny freedom. To achieve this end, we asked all participants to respond to several forced-choice ("Yes"/"No") about freedom, inevitability, and responsibility in two scenarios involving determinism (D) and inevitability (CI; see Scenarios). Depending on their answers to questions at this initial level, we then used two levels of conditional branching to assign participants to complete further materials (see Measures) designed to determine which concept of freedom they used. Those who asserted freedom of action were asked whether
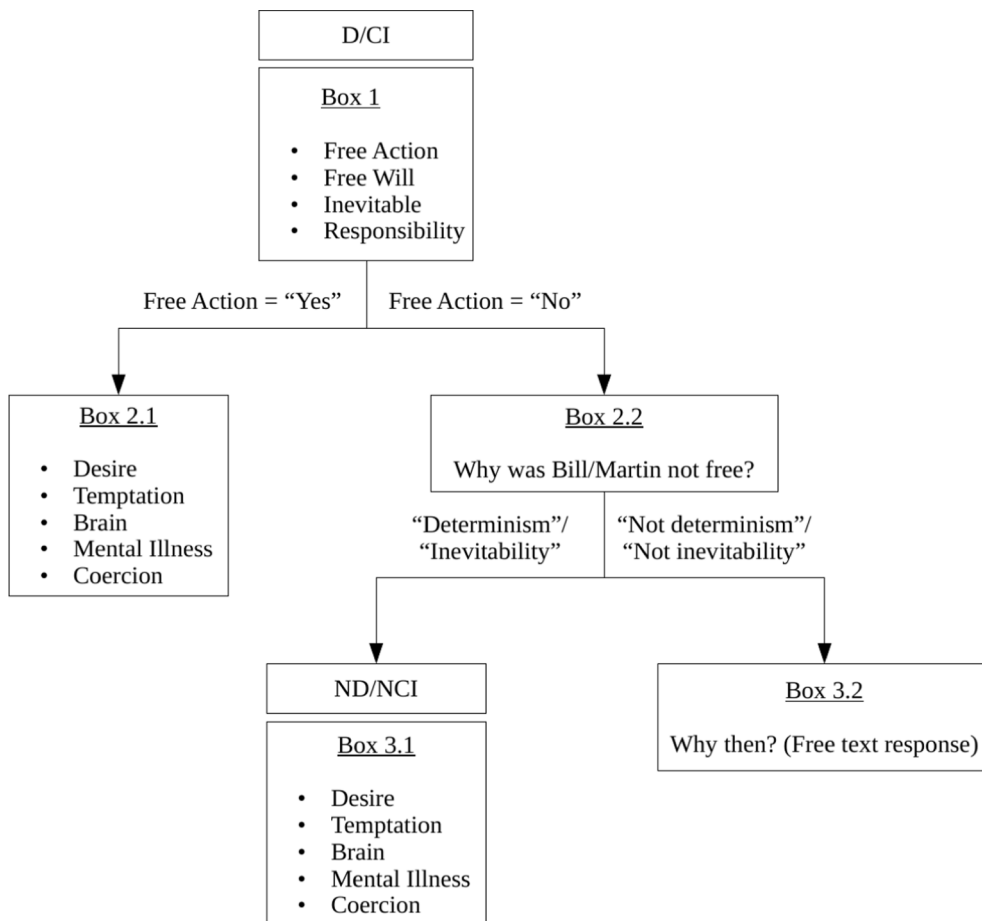

Fig. 1. Structure of Study I.

causation by desire, temptation, normal brain activity, mental illness, or coercion would exclude freedom. Those who denied freedom of action were asked whether they denied this because of determinism in one scenario (D) or inevitability in the other scenario (CI). If they responded negatively, they were asked why they denied freedom. If they responded affirmatively, then they were asked whether causation by desire, temptation, normal brain activity, mental illness, or coercion would exclude freedom in a modified scenario without determinism (ND) or inevitability (NCI). This structure was the same for both scenarios, except that one scenario concerned determinism and the other concerned inevitability. Fig. 1 illustrates this structure. After completing both scenarios, participants answered several self-report questions, and provided demographic information.

2.1.3. Scenarios

All participants read the same two scenarios and responded to questions about them. One of the two scenarios was the "Determined" (D) scenario, which is sometimes called "Actual Sequence" and is frequently used in moral psychology studies (Hannikainen et al., 2019, 3-4; Feltz, Cokely, & Nadelhoffer, 2009; Nichols & Knobe, 2007; Rose & Nichols, 2013). The D scenario describes murder in a deterministic universe:

> *Imagine a universe in which everything that happens is completely brought about by whatever happened before it. This is true from the very beginning of the universe, so what happened in the beginning of the universe brings about what happened next, and so on right up until the present. For example, one day John decided to have vegetable soup at lunch. Like everything else, this decision was completely brought about 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 vegetable soup at lunch.*
>
> *In this universe, a man named Bill has become attracted to his secretary, and he decides that the only way to be with her is to kill his wife and three children. Before he leaves on a business trip, he sets up a bomb that destroys his house and kills his family while he is away.*

The other scenario was the "Counterfactual Intervener" (CI) scenario (Hannikainen et al., 2019, 4; adapted from Frankfurt, 1969; Miller & Feltz, 2011), which describes a murderer who cannot do otherwise:

> *The year is 3072. A group of mad scientists has invented a sophisticated device that can monitor what is going on in a person's mind. The device works at a distance by sending and receiving signals from a special chip that can be easily implanted into a person's brain. With the device, the scientists can change a person's decisions to engage in specific actions by simply sending signals to the special chip implanted in the person's head and thereby manipulating the activation of the person's neurons.*
>
> *One day, the scientists had a person infiltrate a clinic to find people so that the chip could be secretly implanted in them. Martin is one of the subjects who receive the implant.*
>
> *The next day, while monitoring Martin's thoughts, the scientists see that Martin is deliberating on a matter of great concern: whether to kill his friend Adam, who is having an affair with Martin's wife. The scientists agree that they will let Martin make his own decision, but that, if he decides not to kill Adam, they will make him change his mind by sending signals that reinforce his desire and reasons to kill Adam. In other words, regardless of Martin's own final decision, Martin will kill Adam, because the scientists are set on interfering if necessary. Martin decides to kill Adam and ends up killing Adam. The scientists didn't have to interfere.*

Notice that the victim in D (Bill's family) is not guilty, whereas the victim in CI (Adam) might be seen as guilty of having an affair. Because of this difference, we do not rely on any comparison between responses to these different scenarios. Instead, we separately use D to test hypotheses about freedom from determinism and CI to test hypotheses about freedom from inevitability.

2.1.4. Measures

In order to separate participants who affirm freedom from those who deny freedom, we asked them four forced-choice (No = 0/Yes = 1) questions after each of the two scenarios (cf. Fig. 1, Box 1):

- *Free Action.* Did Bill [Martin] act freely when he killed his family [Adam]?
- *Free Will.* Did Bill [Martin] act of his own free will when he killed his family [Adam]?
- *Inevitability.* Was Bill [Martin] able to avoid killing his family [Adam]?
- *Responsibility.* Was Bill [Martin] morally responsible for killing his family [Adam]?

The first question (Free Action) came from Hannikainen et al. (2019). We included separate questions about free action and free will in order to investigate whether subjects distinguished these two concepts and took the same barriers to remove both. We asked about inevitability or ability to do otherwise, because that is supposed to be lacking in the CI scenario and also to test whether determinism is seen as removing ability to do otherwise. Whereas Hannikainen et al. (2019) asked about the blameworthiness of the agents, and how much punishment they deserved, we asked directly about moral responsibility so that we could explore which kinds of freedom are seen as required for moral responsibility.

In addition, participants answered two forced-choice comprehension check questions (No = 0/Yes = 1):

- In Bill's universe, everything that happens is brought about by whatever happened before it. [Martin was able to avoid killing Adam.]
- Bill [Martin] killed his family [Adam].

Answers to these questions should not be affected by participants' views of freedom.

Next, in order to determine which concept of freedom they used, we asked participants questions about which *kinds* of causes they

thought would affect freedom of action in these scenarios. These items were informed by a pilot study (see online supplement, at [Link omitted for anonymous review]). In the pilot, participants answered open-ended questions about D and CI in 10–100 words. We asked which modifications to the scenario would change their response to whether Bill or Martin acted freely or not. The open-ended responses provided by participants informed the phrasing of questions in the studies reported here.

Participants received different variations of the questions based on their answer to whether the agent (Bill or Martin) acted freely (cf. the top branching in Fig. 1). For the D scenario, participants who said that Bill *did* act freely were asked whether they agreed (0 = No/1 = Yes) with a series of conditional statements about the scenario (cf. Fig. 1, Box 2.1):

- *Desire.* If Bill's desire to kill his family caused him to kill his family, then Bill did NOT act freely when he killed his family.
- *Temptation.* If Bill's secretary's beauty caused Bill to kill his family, then Bill did NOT act freely when he killed his family.
- *Brain.* If normal neural activity in Bill's brain caused him to kill his family, then Bill did NOT act freely when he killed his family.
- *Mental Illness.* If Bill had a severe mental illness that caused him to kill his family, then Bill did NOT act freely when he killed his family.
- *Coercion.* If Bill had powerful enemies who threatened to kill him if he did not kill his family, and this threat caused Bill to kill his family, then Bill did NOT act freely when he killed his family.

These conditional statements were informed by open-ended responses collected in our pilot study (see online supplement, at [Link omitted for anonymous review]). The first three causes were chosen to represent causes that are not constraints (Desire, Temptation, Brain), one of which was mental (Desire), another of which was physical (Brain), and the third of which referred to external circumstances (Temptation). The last two causes were chosen to represent causes that are constraints, either internal (Mental Illness) or external (Coercion). Thus, participants who are thinking of freedom from all kinds of causation should respond "No" to all five questions, since they all refer explicitly to causes, whereas participants who are thinking of freedom from constraint should respond "No" to only the last two questions, since only those refer to constraints.

In contrast, participants who said that Bill *did not* act freely in D were first asked whether they agreed (0 = No/1 = Yes) that Bill did not act freely when he killed his family because Bill's act of killing his family was completely brought about by what happened before it (cf. Fig. 1, Box 2.2). Participants who did not agree with this statement were then asked to describe why they thought that Bill did not act freely when he killed his family in 1–2 sentences (cf. Fig. 1, Box 3.2). Participants who did agree with the statement were then shown a modified version of the scenario:

> *Imagine an alternative universe where human decisions are not completely brought about by whatever happened before them. In this universe, a man named Bill has become attracted to his secretary, and he decides that the only way to be with her is to kill his wife and three children. Before he leaves on a business trip, he sets up a bomb that destroys his house and kills his family while he is away.*

We then asked participants whether they agreed (0 = No/1 = Yes) with the same conditional statements as above (Desire, Temptation, Brain, Mental Illness, Coercion; Fig. 1, Box 3.1), but about the modified D (ND) scenario. Because we asked about these statements after modifying the D scenario, we do not rely on any comparison between responses in these two conditions (Boxes 2.1 and 3.1).

For the CI scenario, participants who said that Martin *did* act freely were asked whether they agreed with the same five statements as above (Desire, Temptation, Brain, Mental Illness, Coercion; cf. Fig. 1, Box 2.1) modified to be about Martin killing Adam (so that Martin's temptation to kill Adam was created by Adam's affair with Martin's wife). The modified CI scenario is referred to as NCI (for "No Counterfactual Intervener"). Those who said that Martin *did not* act freely were asked whether they agreed with the statement that Martin was not free because he was unable to avoid killing Adam (cf. Fig. 1, Box 2.2). Participants who disagreed were then asked to describe why they thought that Martin did not act freely in 1–2 sentences (cf. Fig. 1, Box 3.2). Participants who agreed that Martin was not free because he was unable to avoid killing Adam were then shown the following alternative version of the scenario:

Table 1
Proportion of agreement with and Pearson's product-moment correlations between our initial series of questions (cf. Fig. 1, Box 1) in D and CI; square brackets show 95% confidence intervals.

D					
	% Yes	1.	2.	3.	4.
1. Free Action	71.1 [66.0, 75.6]	–	0.92 [0.91, 0.93]	0.56 [0.52, 0.60]	0.64 [0.60, 0.67]
2. Free Will	73.1 [68.2, 77.5]		–	0.55 [0.52, 0.57]	0.65 [0.61, 0.69]
3. Inevitability	47.1 [41.8, 52.4]			–	0.37 [0.34, 0.40]
4. Responsibility	84.5 [80.3, 88.0]				–
CI					
1. Free Action	88.3 [84.5, 91.3]	–	0.90 [0.89, 0.92]	0.15 [0.12, 0.19]	0.69 [0.66, 0.72]
2. Free Will	88.0 [84.1, 91.0]		–	0.18 [0.13, 0.23]	0.71 [0.68, 0.75]
3. Inevitability	22.8 [18.7, 27.5]			–	0.14 [0.14, 0.15]
4. Responsibility	91.2 [87.8, 93.8]				–

In this alternative universe, there are no scientists who put a chip in Martin's brain, so he is able to avoid killing Adam. Martin will kill Adam only if his final decision is to kill Adam. If Martin's final decision is not to kill Adam, then he will not kill Adam. Martin decides to kill Adam and ends up killing Adam.

We then asked these participants whether or not they agreed with the same statements as above (Desire, Temptation, Brain, Mental Illness, Coercion; cf. Fig. 1, Box 3.1), but about the NCI scenario. As before, we do not rely on any comparison between responses after NCI (Box 3.1) and responses after CI (Box 2.1).

Finally, before exiting the survey, participants were asked whether they had encountered either of the scenarios before ("No, neither scenario"; "Yes, the scenario with Bill"; "Yes, the scenario with Martin"; "Yes, both the scenario with Bill and with Martin") and if they had ever taken a philosophy course before ("None"; "1 course"; "2 courses"; ">2 courses"), and provided demographic information.

2.2. Results

All analyses were carried out in R (R Core Team 2020), unless otherwise stated. First, for the initial four questions (cf. Fig. 1, Box 1), we calculated the proportion of participants who agreed with each question. We also calculated pairwise Pearson's product-moment correlation coefficients between participants' responses to these four questions, Table 1 shows the results. The results reported in Tables 1 closely matched previous results in our prior study (see online supplement, at [Link omitted for anonymous review]).

Table 1 shows that in both D and CI, very similar proportions of participants agreed that the agent acted freely and that the agent had free will (items 1 & 2). Moreover, agreement with the two items was highly positively correlated ($rs \geq 0.90$). Only 5.0% [3.1%, 7.8%] of participants gave different answers to these items for at least one of the scenarios (D: 3.2% [1.8%, 5.7%]; CI: 2.0% [1.0%, 4.2%]). Together, this suggests that most participants did not differentiate between the concepts of freedom of action and freedom of will, at least not in D and CI. They still might distinguish free action and free will in other conditions.

Did our participants distinguish avoidability (or non-inevitability) from freedom? It might seem surprising that the correlations between inevitability ratings and free will/action ratings were much lower in the CI scenario ($r = 0.15/0.18$) than in the D scenario ($r = 0.56/0.55$). While this discrepancy could be due to some participants using different senses of "free will" and "free action" in CI and D, we think it might also be an artifact of our choice of scenarios. While free will and free action ratings were quite similar in D and CI, inevitability was openly built into CI but not D, so many more participants had to deny avoidability when they asserted freedom in CI, resulting in the pattern of correlations we observed. In any case, our results show that our participants did tend to distinguish avoidability from freedom.

What about responsibility? While ratings of responsibility and freedom were strongly correlated (Cohen, 1988, 80-81) in both scenarios ($rs \geq 0.60$), a minority of participants (19.3%, [15.5%, 23.8%]) did give different answers in at least one scenario. Of these participants, the vast majority (90.0% [80.8%, 95.1%]) said that the agent did not act freely but was responsible, with the remaining participants (10.0% [4.9%, 19.2%]) ascribing freedom but denying responsibility. Thus, at least these participants must distinguish free action and moral responsibility, and most of them think that an agent can be responsible even if they are not free.

To get a better sense the proportion of people who think that agents can be responsible even if they are not free, we need to look more narrowly at only the participants who denied that Bill acted freely in D because his act was determined. This is because the many participants who said that the agent *was* free were not in a position to give the combined response that the agent was responsible and *not* free. Only those participants who denied freedom were able to ascribe responsibility without freedom. Of the participants who denied that Bill acted freely in D because his act was determined, 49.0% [39.3%, 58.7%] ascribed moral responsibility despite denying free action. Similarly, of the participants who denied that Martin acted freely in CI because his act was inevitable, 45.5% [29.8%, 62.0%] still ascribed moral responsibility anyway. Thus, a large proportion of those who denied freedom and hence were able to ascribe moral responsibility without freedom did indeed ascribe moral responsibility without freedom, though only a small number of participants could take this position because they denied freedom (cf. Table 1).

Next, participants who answered that the agent (Bill or Martin) *did* act freely were asked whether they agreed with a series of conditional statements about the scenario (cf. Fig. 1, Box 2.1). Table 2 lists proportions of agreement for each statement.

Table 2 shows that our participants agreed with Mental Illness and Coercion (the statements about constraint causes) much more frequently than they agreed with Desire, Temptation, and Brain (the statements about non-constraint causes). Because these conditional statements say that a kind of cause makes the agent *not* free, more participants agreed that constraint causes remove freedom than agreed that non-constraint causes remove freedom.

Table 2

Proportions of agreement to conditional statements by participants who answered that the agent did act freely; square brackets show 95% confidence intervals.

	% Yes	
	D ($n = 243$)	CI ($n = 302$)
Desire	6.6 [4.1, 10.4]	4.3 [2.5, 7.2]
Temptation	7.4 [4.7, 11.4]	6.6 [4.3, 10.0]
Brain	15.6 [11.6, 20.7]	11.6 [8.5, 15.7]
Mental Illness	58.0 [51.7, 64.1]	61.3 [55.7, 66.6]
Coercion	47.3 [41.1, 53.6]	54.3 [48.7, 59.8]

In D, almost all participants who indicated agreement with at least one conditional statement agreed with at least one constraint statement (95.7% [91.7%, 97.8%]). Moreover, a sizable majority (73.1% [66.3%, 79.0%]) of participants who agreed with at least one statement indicated agreement with *only* constraint statements. Thus, most people who said that Bill did act freely saw mental illness and coercion as obstacles that remove freedom but did *not* see desire, temptation, or normal brain activity as obstacles to freedom. This supports the idea that a majority of participants did think that Bill was free because they were thinking of freedom as freedom from constraint. In contrast, very few participants (1.6% [0.6%, 4.6%]) agreed with all five of the conditional statements. Indeed, very few (5.4% [2.9%, 9.6%]) agreed with four or more of the statements. This suggests that almost no participants thought of freedom in terms of freedom from (all types of) causation. They also must not have been thinking in terms of freedom from determination because they ascribed freedom in a world that is determined. Together, these findings support (H1).

The picture is similar for CI. Of the participants who agreed with at least one conditional statement, 97.5% [94.7%, 98.9%] agreed with at least one constraint statement, and 80.3% [74.9%, 84.8%] agreed with *only* constraint statements. Moreover, very few participants (2.0% [0.9%, 4.7%]) agreed with all five of the conditional statements. Indeed, very few (3.7% [2.0%, 6.9%]) agreed with four or more of the statements. Again, this suggests that these participants were ascribing freedom to Martin because they were thinking of freedom from constraint, not freedom from causation. They also must not have been thinking in terms of freedom from inevitability because they ascribed freedom to an agent who could not do otherwise. Together, these two findings support (H3).

Of participants who answered that Bill did *not* act freely in D, all but one ($n = 98$; 99.0% [94.5%, 99.9%]) agreed to the statement "Bill did not act freely when he killed his family because Bill's act of killing his family was completely brought about by what happened before it" (cf. Fig. 1, Box 2.2). Of the participants who answered that Martin did *not* act freely in CI, $n = 33$ (82.5% [68.1%, 91.3%]) agreed with the statement "Martin did not act freely when he killed Adam because he was not able to avoid killing Adam." These findings support the positive parts (a) of (H2) and (H4)—that is, they suggest that most participants who denied freedom in D were thinking of freedom from determination, and most participants who denied freedom in CI were thinking of freedom from inevitability (or inability to do otherwise). Perhaps those who denied freedom in D might have been thinking of freedom from inevitability and inferred inevitability from determinism. Our findings of mediation by Inevitability in Study 2 below could suggest this possibility. Nonetheless, what our subjects said was that Bill's act was not free because it was determined, so the simplest and most natural interpretation is that they were thinking of freedom from determination.

Next, participants who agreed with these statements about why Bill or Martin was not free, read and responded to the ND and NCI (cf. Fig. 1, Box 3.1). Table 3 lists proportions of agreement for each conditional statement.

Table 3 shows that our participants in this group agreed with Mental Illness and Coercion (the statements about causes that are constraints) much more frequently than they agreed with Desire, Temptation, and Brain (the statements about non-constraint causes). Moreover, in ND, almost all participants who indicated agreement with at least one statement agreed with at least one constraint statement (95.1%, [88.0%, 98.1%]). Moreover, a large majority (70.4% [59.7%, 79.2%]) of these participants agreed with *only* constraint statements. Similarly, in our NCI scenario, of the participants who agreed with at least one conditional statement, 86.7% [70.3%, 94.7%] agreed with at least one constraint statement, and 60.0% [42.3%, 75.4%] agreed with *only* constraint statements. These results suggest that most participants who denied that the agent acted freely in the original scenarios were thinking about freedom from constraint at the time when they were asked whether they agreed with our five statements about the modified scenarios, ND and NCI.

Regarding freedom from causation, few of our participants who were asked about our ND agreed with all five statements (9.9% [5.1%, 18.3%]) or with at least four out of five statements (12.3% [6.8%, 21.3%]). Similarly, few of our participants who were asked about our NCI agreed with all five statements (10.0% [3.5%, 25.6%]) or even with at least four out of five statements (13.3% [5.3%, 29.7%]) in our ND. The percentages for ND and NCI are larger than those for the original D and CI. However, only one of these differences is statistically significant: The proportion of participants who responded to ND and agreed with all five conditional statements was larger than the proportion of participants who responded to the original version of D and agreed with all five conditional statements, $\chi^2(1) = 7.775$, $p = .005$ (all other χ^2s < 3.58, ps > 0.059). Hence, while this may suggest that more participants who said that Bill was not free in D thought of freedom of causation, our results should not be over-interpreted in this regard.

2.2.1. Principal components analyses

Our overall method relies on the assumption (frequently made by philosophers and psychologists, e.g. Dennett, 1983; Monroe & Malle, 2010) that there is a meaningful distinction to be drawn between causes that are constraints (including Mental Illness and Coercion) and causes that are not constraints (including Desire, Temptation, and Brain). To check this assumption, we subjected our

Table 3
Proportions of agreement to conditional statements by participants who read the ND and NCI; square brackets show 95% confidence intervals.

	% Yes	
	ND ($n = 98$)	NCI ($n = 33$)
Desire	12.2 [7.1, 20.2]	21.2 [10.7, 37.8]
Temptation	13.3 [7.9, 21.4]	15.2 [6.7, 30.9]
Brain	19.4 [12.8, 28.3]	24.2 [12.8, 41]
Mental Illness	66.3 [56.5, 74.9]	60.6 [43.7, 75.3]
Coercion	59.2 [49.3, 68.4]	60.6 [43.7, 75.3]

conditional statements to a series of principal component analyses (PCA). A PCA clusters linear combinations of original variables by creating new orthogonal implicit variables to reduce the dimensionality of a dataset and account for the maximum amount of variance (Jolliffe & Jorge, 2016). Factor loadings represent an item's correlation with the extracted component. We used JASP (JASP Team, 2020) for this purpose.

PCAs applying promax rotation for D, CI, ND, and NCI all resulted in a two-component fit (for details, see the online supplement: [Link omitted for anonymous review]). Tables 4 and 5 show the results. Desire, Temptation, and Brain always strongly loaded on one component (I), Mental Illness and Coercion on the other (II). Since the items loaded onto orthogonal implicit components for each scenario, this suggests that participants indeed distinguished between conditional statements about causes that are constraints and conditional statements that about causes that are not constraints.

A principal component analysis that collapsed across the D, ND, CI, and NCI scenarios also found a distinct component for the first three statements (about causes that are not constraints) but further subdivided the response for the last two statements (about causes that are also constraints) into two separate components (here labelled IIa and IIb). Specifically, severe mental illness and coercion loaded separately. We interpret these three components to represent: (I) causes that are not constraints, (IIa) causes that are also internal constraints (exemplified by mental illness), and (IIb) causes that are also external constraints (exemplified by coercive threats). However, another possible interpretation refers to the distinction in law and morality between justifications (which show an act is not wrong) and excuses (which admit an act is wrong but show its agent is not responsible). Component (IIa) represents constraints (like mental illness) that are often seen as excuses but not justifications, whereas component (IIb) represents constraints (like coercion) that are often seen as justifications instead of excuses (since giving a thief your money in order to save your life is not wrong). In any case, this analysis should not be given too much weight, because it compares across four scenarios that differ in relevant respects, including whether determinism holds and whether the victim was guilty of prior misbehavior (in CI but not D; see above).

2.3. Discussion

Our main goal in Study I was to probe which concepts of freedom various people use when they ascribe or deny freedom of action. Participants read two scenarios, one about an agent in a deterministic universe who murders his family (D), one about an agent who lacks the ability to do otherwise who murders his friend (CI). Participants who affirmed that the agent *did* act freely were then asked whether they agreed with five conditional statements about whether the act would not have been free if it had been caused by different kinds of factors, including an internal (mental illness) or external (coercion) constraint on the agent or a cause that was not commonly recognized as a constraint (desire, temptation, or normal neural activity). Participants who instead responded that the agent *did not* act freely because the act was determined (in D) or inevitable (in CI) were shown a modified version of the scenario in which the universe was not determined (in the ND) or the agent was able to do otherwise (in the NCI), and then they indicated whether they agreed with the same series of conditional statements.

Our results confirm some but not all of our hypotheses. Regarding (H1), a large majority of participants who agreed with at least one conditional statement about the D scenario agreed only with conditional statements that described an internal or external constraint on the agent. This indicates that most participants thought that only causes that are constraints (like mental illness and coercion), but not other kinds of causes (like desire, temptation, or neural activity), remove freedom. Since the act in D is not constrained, this finding suggests that these participants were thinking in terms of freedom from *constraint*. Moreover, they must *not* have been thinking about freedom from *determination* in D, because the scenario specified explicitly that the agent was determined in D, and these participants passed comprehension checks. Thus, our results supported our hypothesis (H1). A similar pattern emerged for responses to the CI scenario, which supports (H3).

The positive parts (a) of (H2) and (H4) were also supported. Participants who denied that the agent in D was free were asked whether this was because the act had taken place in a universe in which everything that happens is brought about by whatever happened before it (that is, a deterministic universe). All but one answered "Yes." This near unanimity suggests that these participants were thinking about freedom from *determination*, as predicted by part (a) in (H2). Similarly, a large majority of participants who denied freedom in CI indicated that this was because the agent lacked the ability to do otherwise. This result suggests that these participants were thinking about freedom from *inevitability*, which confirms (a) in (H4).

In contrast, we did not find support for the negative parts (b) of (H2) and (H4). Of the participants who read the ND scenario and who agreed with at least one conditional statement, a large majority agreed only with statements that described constraints on the agent. This indicates that most of these participants thought that only constraints (like mental illness and coercion), but not other kinds of causes (like desire, temptation, or neural activity), remove freedom in ND. A similar pattern emerged for responses on the NCI

Table 4
PCA using promax rotation of responses to the conditional statements for D, CI, ND and NCI: Component loadings (two-component solution).

	D		ND		CI		NCI	
	I	II	I	II	I	II	I	II
Desire	0.771	–	0.790	–	0.764	–	0.906	–
Temptation	0.764	–	0.923	–	0.795	–	0.923	–
Brain	0.729	–	0.860	–	0.716	–	0.556	–
Mental Illness	–	0.779	–	0.869	–	0.785	–	0.809
Coercion	–	0.758	–	0.755	–	0.725	–	0.751

Table 5
PCA using promax rotation of responses to the conditional statements for D, CI, ND and NCI: Component characteristics (two-component solution).

	EV		% of variance		Cumulative %	
	I	II	I	II	I	II
D	1.753	1.152	0.351	0.230	0.531	0.581
ND	2.403	1.163	0.481	0.233	0.481	0.713
CI	1.795	1.085	0.359	0.217	0.359	0.576
NCI	2.041	1.205	0.408	0.241	0.649	0.649

scenario. The negative parts of (H2) and (H4) hypothesized that people are not thinking about freedom from constraint when they deny that an agent acts freely because the act is determined or inevitable. Our results did not support these hypotheses.

Finally, few participants who denied freedom in D and CI indicated agreement with four or more of our five conditional statements. This suggests that few (if any) participants think that all causes, including causes that are not constraints (like desires, temptations, and normal neural activity) are barriers to freedom. In other words, our study suggests that very few people think of freedom of action as freedom from causation. In contrast, some philosophers (e.g. Ginet, 1997; McCann, 2012) take causation to remove freedom of action. The philosophers still might be correct, but the absence of this concept of freedom among our participants suggests that these philosophers are not using the same concept as lay persons.

In addition to our main hypotheses about what lay persons think of freedom of action as freedom from—that is, what they contrast freedom with—we also wanted to investigate whether they distinguish freedom of action from freedom of the will. For both scenarios, responses to our questions about whether the agent (Bill or Martin) acted freely or acted of his own free will very strongly correlated, and almost no participants gave different answers to the two questions. This suggests that most people do not make a distinction between freedom of the will and free action, at least not in our scenarios. This assimilation is especially striking in CI because many philosophers claim that Martin could have free will without free action (Gert & Duggan, 1979).

Our other subsidiary goal was to explore the relation of responsibility to freedom. Around half of participants who denied that the agent acted freely because of determinism in D or inevitability in CI still said that the agent was responsible. The small number of participants who denied that either agent acted freely makes this finding insecure, though it aligns with some prior studies (Figdor & Phelan, 2015; Vierkant et al., 2019). Anyway, at least this minority of our participants ascribed responsibility without freedom, perhaps because they used a forward-looking concept of responsibility or traced responsibility back to prior free actions. The rest of our participants ascribe or deny both freedom and responsibility together in these scenarios, but they still might distinguish responsibility from freedom and might even ascribe responsibility without freedom in other scenarios. However, our results here do not address this possibility.

Our results are in line with previous research on the lay concept of free will in several ways. First, our study supports the view that most people think of free action as requiring freedom from constraint, not from causation (Monroe, Dillon, & Malle, 2014; Monroe & Malle, 2010; Stillman, Baumeister, & Mele, 2011; Vonasch, Baumeister, & Mele, 2018). Regardless of whether participants said that the agent in D or CI acted freely, a large majority of them seemed to be operating at times with a notion of freedom as freedom from constraint, but not freedom from causation.

Second, our results also support the view that a significant proportion (29%) of people think of freedom of action as removed by determinism (cf. Nichols and Knobe, 2007; Sarkissian et al., 2010). Nonetheless, as mentioned above, they still might be thinking about freedom from constraint if they mistakenly assume that determinism rules out causation by mental states (beliefs, desires, decisions, plans, etc.) so that every act in a deterministic world is constrained (Murray & Nahmias, 2014; Nahmias & Murray, 2011). Our next study explores this possibility.

3. Study II

In Study I, we found that people who deny freedom in D and CI largely seem to think about freedom from constraint when they are later considering our modified versions, ND and NCI. This result is susceptible to (at least) two interpretations.

On the first interpretation, this subset of our participants might have been assuming all along that determination and inevitability remove freedom *because* determination and inevitability introduce constraints on the agent. They might think this because determination and inevitability imply that the agent's desires do not cause the action, so the causal chain leading to the action bypasses the agent's mental states (including desires and decisions). Although Nahmias and Murray (2011) developed their bypassing hypothesis as an error theory to explain why participants denied freedom in the abstract case in Nichols and Knobe (2007), we will instead investigate participants who deny freedom in our more concrete scenarios. The *Bypassing Hypothesis* that concerns us claims that people who hold that determinism is incompatible with freedom believe so because they assume that determinism excludes mental causation. If so, these participants might have been thinking about freedom from constraint even while they judged the agents not to act freely in the original D and CI scenarios.

On the second interpretation, participants might instead think about freedom from determination and inevitability at the time when they judged that the agent did not act freely in the original D and CI scenarios, where the act was determined or inevitable. They might then have switched to a different concept of freedom—freedom from constraint—after they changed from the original scenarios to our modified scenarios (ND and NCI), where the act was *not* determined or inevitable. In other words, they might think about freedom from constraint when determinism and inevitability are out of the picture, but not think about freedom from constraint when

they are considering deterministic universes and counterfactual interveners who make acts inevitable. This interpretation can be described as a *Contextual Hypothesis* because it claims that people use different concepts of freedom in deterministic contexts than they use in indeterministic contexts.

Study II was designed to test the Bypassing Hypothesis. If the Bypassing Hypothesis is ruled out by the evidence, that provides some evidence for the Contextual Hypothesis, at least in the absence of any other plausible interpretation of our findings in Study I. Study II was pre-registered at: [Link omitted for anonymous review].

3.1. Materials and methods

3.1.1. Participants

Participants were recruited through the online subject pool Prolific (https://www.prolific.co.uk). We restricted participation to US residents whose first language was English, had completed at least 50 other studies in Prolific, and had at least a 90% acceptance rating on Prolific.3.1 Participants were compensated for their time, and our study was approved by the campus IRB of Duke University.

The sample size was determined prior to data collection via power analysis. In a pilot sample, we found an *ICC* = 0.22 for participant's responses to Freedom. According to Pan et al. (2018, Table 2), this meant that we would require a sample size of $n > 340$ to detect small mediation effects with power of at least 80%. In order to account for the exclusion of participants due to comprehension check failures, we collected a sample of $n = 445$ (66.1% Female, 33.3% Male, 0.4% Other, 0.2% Prefer not to answer). 134 participants failed at least one comprehension check (see Measures) and were therefore excluded from all analyses. (Apart from two instances which we describe when we get to them, none of our reported results change in a substantial way if these participants are included.) This leaves us with a final sample of $n = 311$ (63.7% Female, 35.4% Male, 0.6% Other, 0.3% Prefer not to answer).

3.1.2. Design

Our study used a within-subject design with two conditions (Determined/Not-Determined). Each participant read and responded to a series of questions about two scenarios, in random order. Participants then provided demographic information and exited the survey.

3.1.3. Scenarios

All participants read two scenarios in randomized order. The scenarios had agent names randomized between Bill and Martin. In the Determined condition, participants read the D scenario from Study 1. In the Not-Determined condition, participants read a modified version of the ND scenario from Study 1, with three additional sentences inserted after the first sentence to make it clearer to participants that acts in this universe are not determined:

> Imagine an alternative universe where human decisions are not completely brought about by whatever happened before them. For example, one day [agent] decided to have vegetable soup at lunch. Like other decisions, this act was not completely brought about by what happened before it. So, if everything in this universe was exactly the same up until [agent] made his decision, then [agent] still might have instead decided not to have vegetable soup at lunch. In this universe, a man named Bill has become attracted to his secretary, and he decides that the only way to be with her is to kill his wife and three children. Before he leaves on a business trip, he sets up a bomb that destroys his house and kills his family while he is away.

3.1.4. Measures

For each scenario, we measured participants' level of agreement with the following statements (all continuous scales, from "Completely disagree" [$= -100$] to "Completely agree" [$= 100$]):

- *Freedom.* [Agent] acted freely when he killed his family.
- *Responsibility.* [Agent] is morally responsible for killing his family.
- *Inevitability.* [Agent] was able to avoid killing his family.
- *Desire.* [Agent]'s desire to be with his secretary caused Bill to kill his family.
- *Decision.* [Agent]'s decision to kill his family caused [agent] to kill his family.
- *Neural Causation.* Neural activity in [agent]'s brain caused [agent] to kill his family.

For each scenario, we also asked participants to answer two forced-choice comprehension check questions ("Yes"/"No"):

- In [agent]'s universe, everything that happens is brought about by whatever happened before it.
- [Agent] killed his family.

Finally, after completing both scenarios, participants were asked to rate on a scale from "completely implausible" ($=-100$) to "completely plausible" ($=100$) the plausibility of a universe in which everything that happens is completely brought about by whatever happened before it (that is, a deterministic universe like the one described in D) and a universe where human decisions and actions are not completely brought about by whatever happened before them (that is, an indeterministic universe like the one described in ND).

3.2. Results

All analyses were carried out in R (R Core Team 2020). We begin by calculating means, standard deviations, and pair-wise Pearson's product-moment correlation coefficients for our main measured variables (Free, Responsible, Decision, Desire, Neural). Table 6 shows these results.

In order to investigate whether participants in this study responded to Freedom (called Free Action in Study I), Responsibility, and Inevitability in a similar way as the participants in Study I, we dichotomized participants' responses to these three questions (response $\leq 0 \rightarrow$ "No"; response $> 0 \rightarrow$ "Yes") and compared the proportions of "Yes" responses to the proportions we observed in the previous study using two-proportions z-tests (with Yates' continuity correction). The proportions did not differ significantly for any of the three questions (Freedom: 65.3% vs. 71.1% in Study I; Responsibility: 86.8% vs. 84.5%; Inevitability: 53.7% vs. 47.1%; $ps > 0.106$, $\chi^2 < 2.60$).[1] This analysis, along with all following analyses (unless otherwise stated), was pre-registered at [Link omitted for anonymous review].

To investigate the Bypassing Hypothesis, we used a pair of model-based within-subject mediation analyses (Kenny & Kenny, 2003; Vuorre & Bolger, 2018). According to the Bypassing Hypothesis, people do not believe that determinism in and of itself precludes freedom. Instead, when they hold that determinism is incompatible with freedom, this is because they assume that determinism excludes mental causation, which in turn is a necessary condition for freedom. If so, then participants' responses to Desire and Decision (both statements about mental causation) should mediate the relationship between whether the act took place in a deterministic universe or not and their ratings of whether the agent acted freely. Fig. 2 shows the corresponding pair of mediation models.

All mediation analyses reported in this paper were carried out using the *mediation* package (Tingley et al., 2014), which takes two model objects as input, the mediator model (the conditional distribution of the mediator given the treatment—in our case, whether the act took place in a deterministic universe or not) and the outcome model (the conditional distribution of the outcome given both the mediator and the treatment). Models were fit using ML (Bates et al., 2015); for each model, we specified a random intercept for participant. All proposed mediating variables were standardized, to allow for easier interpretation of total effects. For each proposed mediating variable, we estimated the average causal mediation effect (*ACME*) with quasi-Bayesian 95% confidence intervals (number of simulations = 100). In our pre-registration, we had planned to use bootstrap 95% CIs. However, this turned out not to be possible with the *mediation* package, which is why we used quasi-Bayesian 95% confidence intervals instead.

An anonymous reviewer argues that participants often interpret items like Desire and Decision to mean that the desire or decision in question is something like an irresistible force or compulsion. For example, the reviewer suggests that many participants may have read Desire as "[Agent] was overwhelmed by a strong desire to kill his family that compelled him to kill his family even if that was contrary to some of his other beliefs". In support, the reviewer points to the low rates of "Yes" responses to Desire (and to a lesser extent, Decision) even in ND. If this is true, then our mediation analysis does not test the bypassing hypothesis because the bypassing hypothesis is not about desires and decisions that are irresistible to the agent.

To investigate this worry, we re-ran Study II (Study II*) with revised versions of Desire and Decision:

- *Desire**. [Agent] killed his family because he desired to be with his secretary.
- *Decision**. [Agent] killed his family because he decided to kill his family.

Rewording in terms of "because" was suggested by the anonymous reviewer as a way to avoid the misinterpretation that the reviewer was worried about. Except for these two changes, Study II* used the same design, scenarios and measures as before. Details about participants, as well as means, standard deviations and pair-wise Pearson's product-moment correlation coefficients for the main measured variables can be found online at [Link omitted for anonymous review]. In what follows, for each mediation model that includes Desire or Decision, we also report results for the same model using Desire* or Decision* instead (not pre-registered). Compared to Desire (see, Table 6), mean ratings for Desire* were much higher in both scenarios (D: $M = 27.8$ vs. 74.0; ND: $M = 30.9$ vs. 76.0). In contrast, mean ratings for Decision and Decision* were very similar (D: $M = 51.1$ vs. 49.4; ND: $M = 78.1$ vs. 79.5).

There was a strong total effect of the experimental condition on responses to Freedom, total effect $= -0.92$ $[-1.08, -0.79]$ (Study II*: total effect $= -0.82$ $[-0.96, -0.69]$), such that participants ascribed less freedom to the agent in D than to the agent in ND. We found that Desire did not mediate this effect, *ACME* $= 0.00$ $[-0.02, 0.02]$ (Desire*: *ACME* $= -0.01$ $[-0.04, 0.02]$) (the confidence interval includes 0). In contrast, responses to Decision did partially mediate this effect, *ACME* $= -0.15$ $[-0.20, -0.10]$ (Decision*: *ACME* $= -0.21$ $[-0.27, -0.16]$). In order to gauge the magnitude of this partial mediation effect, we calculated the proportion of the total effect that is mediated, *pme* $= ACME/$total effect (Vuorre & Bolger, 2018, p. 2137–8). We find that the mediation effect of responses to Decision was only weak, *pme* $= 0.16$ (Decision*: *pme* $= 0.26$). Together, these results do not support the Bypassing Hypothesis, no matter which wording of Decision and Desire we use.

The Bypassing Hypothesis claims that people who hold that determinism is incompatible with freedom believe so because they assume that determinism excludes mental causation. There are two ways to understand what it means for people to hold that determinism is incompatible with freedom. On one view, people hold that determinism is incompatible (to some degree) with freedom if they ascribe less freedom in a deterministic universe than in an indeterministic universe, even if they affirm some freedom in both

---

[1] If we do not exclude participants who failed at least one comprehension check, then the proportions of "Yes" responses for Responsibility and Inevitability do differ significantly between Study II and Study I (Responsibility: 89.7% vs. 84.5%, $p = 0.040$, $\chi^2 = 4.22$; Inevitability: 61.8% vs. 47.1%, $p < 0.001$, $\chi^2 = 16.4$).

C. Simmons et al.

Consciousness and Cognition 101 (2022) 103318

Table 6
Means, standard deviations and pair-wise Pearson's product-moment correlation coefficients for our main measured variables.

|	D							
	M	SD	1.	2.	3.	4.	5.	6.
1. Freedom	26.1	72.7	–	0.65 [0.59, 0.71]	0.64 [0.58, 0.70]	0.18 [0.10, 0.25]	0.49 [0.42, 0.56]	0.14 [0.07, 0.14]
2. Responsibility	63.5	59.1		–	0.52 [0.44, 0.59]	0.32 [0.24, 0.40]	0.55 [0.48, 0.62]	0.16 [0.09, 0.23]
3. Inevitability	12.1	81.9			–	0.09 [0.05, 0.13]	0.39 [0.31, 0.47]	0.11 [0.05, 0.16]
4. Desire	27.8	76.4				–	0.23 [0.15, 0.31]	0.04 [0.04, 0.05]
5. Decision	51.1	65.6					–	0.28 [0.19, 0.36]
6. Neural Causation	3.3	68.7						–
	ND							
	M	SD	1.	2.	3.	4.	5.	6.
1. Freedom	84.0	28.0	–	0.34 [0.25, 0.43]	0.34 [0.24, 0.42]	0.1 [0.02, 0.18]	0.25 [0.15, 0.34]	0.09 [0.02, 0.17]
2. Responsibility	91.7	22.3		–	0.23 [0.14, 0.32]	0.08 [0.03, 0.14]	0.24 [0.14, 0.33]	0.12 [0.03, 0.21]
3. Inevitability	69.9	57.3			–	−0.04 [−0.05, −0.04]	0.21 [0.12, 0.3]	0.08 [0.02, 0.15]
4. Desire	30.9	74.7				–	0.11 [0.03, 0.19]	−0.05 [−0.09, −0.01]
5. Decision	78.1	42.2					–	0.15 [0.06, 0.24]
6. Neural Causation	20.1	64.5						–
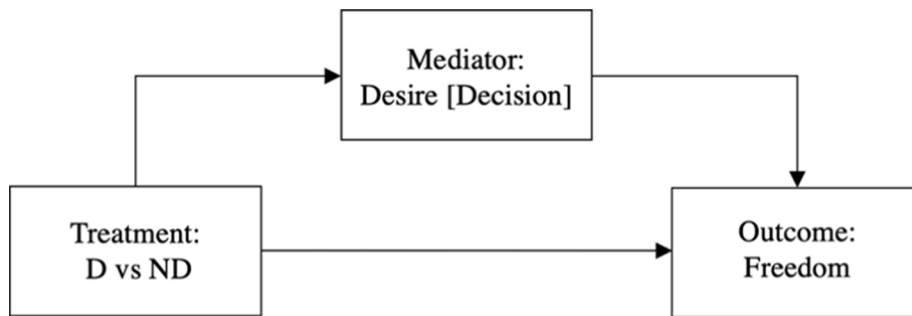
Fig. 2. Mediation models used to test the Bypassing hypothesis.

cases. In order to test the Bypassing Hypothesis on this understanding, we need to determine whether the fact that our participants ascribed less freedom in D than in ND can be explained by their responses to our mental causation statements. This is what the mediation analysis presented in the last two paragraphs did.

In contrast, many philosophers would deny that incompatibility comes in degrees. On their view, people believe that determinism is incompatible with freedom only if they deny freedom in a determined universe. If so, the mediation analysis reported in the last two paragraphs does not test the Bypassing Hypothesis, strictly speaking, because it includes data from participants who affirmed Freedom in D by responding > 0. To test the Bypassing Hypothesis using this dichotomous interpretation of "incompatible", we performed an additional exploratory analysis and ran the same pair of within-subject mediation analyses as before, but restricted to participants who did not ascribe Freedom in D (that is, responded ≤ 0). This analysis was not pre-registered. Again, there was a strong total effect of the experimental condition on responses to Freedom, total effect = −1.80 [−1.92, −1.66] (Study II*: total effect = −1.78 [−1.89, −1.66]), such that participants ascribed less freedom to the agent in D than to the agent in ND. (Note that this is not an independent analysis because participants were selected for freedom ratings in D ≤ 0.) As in the full sample, Desire did not mediate this relationship, *ACME* = −0.01 [−0.03, 0.01] (Desire*: *ACME* = −0.03 [−0.07, 0.00]). In contrast to the full sample, responses to Decision also did not mediate the relationship, *ACME* = −0.05 [−0.11, 0.01]. Moreover, of the 108 participants who denied Freedom in D (that is, responded ≤ 0), only 26.0% [18.6%, 35.0%] also denied mental causation by denying both Desire and Decision (that is, by responding ≤ 0 to both items). This is much less than would be predicted by the Bypassing Hypothesis, which claims that those who deny freedom because of determinism do so because they think that determinism excludes mental causation. And while in Study II*, responses to Decision* did mediate the relationship, *ACME* = −0.16 [−0.25, −0.08], the effect was again only weak, *pme* = 0.13. Thus, the Bypassing Hypothesis was again not supported on a dichotomous reading of "incompatible" (for neither wording of Desire and Decision).

A possible alternative to the Bypassing Hypothesis is that people who hold that determination—or a deterministic universe—is incompatible with free will believe so because they assume that determinism implies that the act was inevitable. To explore this possibility, we conducted an additional within-subject mediation analysis investigating the impact of responses to Inevitability on the relationship between our experimental condition and responses to Freedom. We found that Inevitability did indeed partially mediate this relationship, *ACME* = −0.35 [−0.43, −0.27]. Moreover, this mediation effect was moderate in size (*pme* = 0.38). This suggests that many people's assumption that acts in deterministic universes are inevitable might have an important role to play in their view that freedom is incompatible with determinism.

Again, we asked whether this mediation by Inevitability also holds on the stricter understanding of "incompatible". This analysis

was not pre-registered. If we restrict the mediation analysis to participants who denied freedom in D (that is, responded ≤ 0), we still find that responses to Inevitability mediated the relationship between the experimental condition and ratings of Freedom, $ACME = -0.28$ $[-0.39, -0.16]$. However, in contrast to what we found in the full sample, this effect was only small ($pme = 0.16$). Thus, if we presuppose the dichotomous understanding of "incompatible", then our results do not provide strong support for an important role of people's assumption that acts in deterministic universes are inevitable when they deny freedom in such universes, but not in an indeterministic universe.

Finally, in addition to its (in)compatibility with free will, philosophers often discuss whether agents can be morally responsible for their actions in a deterministic universe. Looking at Table 6, we find a positive mean rating of Responsibility about the agent in D, suggesting that, overall, participants tended to think that this agent was morally responsible for killing his family. On the other hand, the mean rating of Responsibility about the agent in D was considerably lower than the one about the agent in ND ($M = 63.5$ vs. $M = 91.7$). This suggests that participants did seem to think that a deterministic universe removes or diminishes moral responsibility, at least to some degree. Thus, we can ask the question of what (if any) impact perceived mental causation and inevitability might have on this relationship between the experimental condition and responses to Responsibility. To explore this, we conducted additional within-subject mediation analyses investigating the impact of responses to Desire and Decision (both statements about mental causation) and responses to Inevitability on the relationship between our experimental condition and Responsibility.

There was a moderate total effect of the experimental condition on responses to Responsible, total effect $= -0.60$ $[-0.73, -0.47]$ (Study II*: -0.50 $[-0.64, -0.36]$), such that participants ascribed less responsibility to the agent in D than to the agent in ND. We find that responses to Desire did not mediate this effect, $ACME = -0.01$ $[-0.04, 0.03]$ (Desire*: $ACME = -0.02$ $[-0.05, 0.03]$). In contrast, responses to Decision did partially mediate the effect, $ACME = -0.18$ $[-0.29, -0.13]$ (Decision*: -0.21 $[-0.26, -0.16]$); this mediation effect was moderate, $pme = 0.30$ (Decision*: $pme = 0.41$). This suggests that the idea that mental causation *in general* is bypassed in a deterministic universe cannot explain the fact that participants ascribe less moral responsibility in a deterministic universe than in an indeterministic universe. However, believing that certain types of mental causation (for example, an agent's decision-making) are being bypassed in a deterministic universe (or, at least, that their causal importance is diminished) may play some explanatory role.

Finally, we found that responses to Inevitability strongly mediated the relationship between our experimental condition and responses to Responsibility ($ACME = -0.33$ $[-0.41, -0.24]$; $pme = 0.53$). Thus, as for ascriptions of freedom discussed above, the belief that acts are more difficult to avoid in a deterministic universe than in an indeterministic universe likely explains why many participants gave lower ratings of responsibility in D than in ND.

3.3. Discussion

Study II found that participants who took determinism to exclude freedom and responsibility generally did not deny causation by mental states, here represented by desires and decisions. Their ratings of agreement with claims about causation by desires and decisions at most weakly mediated the relation between determinism and freedom or responsibility among this subgroup of our participants.

These results speak against the Bypassing Hypothesis, which claims that people who hold that determinism is incompatible with freedom and responsibility believe so because they assume that determinism excludes mental causation. If our participants believed this, then they would have denied causation by desires and decisions in the Determined Scenario. Their ratings of agreement with claims of mental causation (by desire and decision) would also have strongly mediated the relation between determinism and freedom or responsibility. However, they ascribed mental causation in a deterministic universe, and their views on mental causation did not affect their views on the relation between determinism and freedom or responsibility. Thus, the Bypassing Hypothesis cannot explain why they denied freedom and responsibility in the deterministic universe.

Additional evidence against the Bypassing Hypothesis comes from our analysis of relations to neural causation. Similar results held for responsibility, so, in common parlance, these participants would not accept the excuse "My brain made me do it." Thus, these participants do not seem to follow this line of reasoning for the Bypassing Hypothesis.

Instead of mental and neural causation, we found that inevitability—the lack of ability to do otherwise—did mediate the relation between determination and freedom. This result suggests that people who believe that determinism is incompatible with freedom and responsibility believe so because they assume that determinism makes acts inevitable. Our finding that inevitability works differently than mental causation suggests that participants who think that an agent cannot avoid doing an act do not think so simply because they assume that the act is not caused by the agent's mental states or that the act is caused by the agent's brain. However, these mediation results for inevitability seem to conflict with our result in Study 1 that many participants called the agent free in CI when they also said that the act was inevitable. Further research is needed to explore this apparent conflict, so we cannot yet be confident that inevitability really does mediate the relation between determination and freedom.

Because our results speak against the Bypassing Hypothesis, they also speak in favor of the competing Contextual Hypothesis. Participants in Study I who deny freedom in a deterministic universe (in D) largely seem to think about freedom from constraint when they are later considering our non-deterministic universe scenario (ND). The Contextual Hypothesis explains this pattern by postulating that these participants thought about freedom from determination (and/or inevitability) when determinism was explicitly affirmed but instead thought about freedom from constraint in when determinism was explicitly denied. In the absence of any better explanation of this pattern of responses, these results support the Contextual Hypothesis.

Moreover, these results are also compatible with the negative part of our original hypothesis:

(H2) People who *deny* that an agent acts freely in a scenario where the act is determined (D) are mostly thinking … (b) not about freedom from constraint.

This denial conflicts with the Bypassing Hypothesis but is compatible with the Contextual Hypothesis. Thus, because our results support the Contextual Hypothesis, they do not undermine (H2). (H2) is still not proven, but it is not refuted by our evidence.

The Contextual Hypothesis also fits well with our more general point that different people use different concepts at different times in debates about freedom and responsibility. Our results show how several distinct concepts of freedom—including freedom from constraint, freedom from determination, and freedom from inevitability—all play some role in some lay judgments about free action in some scenarios. That might explain how competing positions on freedom can each receive some support from lay persons.

As a result, we cannot make progress on these philosophical issues or in the corresponding psychological research without separating the different kinds of freedom that different people think about. We need to move beyond simple questions like, "Did Bill act freely?" and instead distinguish different concepts of freedom by deploying more complex experimental designs that ask which changes affect who is free.

4. Limitations and future research

As noted, responses by our sample differed from the data collected by Hannikainen et al. 2019, especially insofar as fewer of their participants said that the agent acted freely in D (47% vs. 71% in our Study I). Further work is needed to explain this variation. One possibility is that we used only W.E.I.R.D. (Western Educated Industrialized Rich Democratic) participants (Henrich et al. 2010), whereas Hannikainen et al. 2019 used a much larger and more diverse sample. Further research should test whether our results replicate in more diverse samples.

Additionally, a subset of the conclusions we draw from Study I are based on analyses that were restricted to participants who denied freedom of action in D and CI. However, these sub-samples were small, meaning that any conclusions we draw from them need to be viewed with caution. Therefore, future research might seek larger samples.

Finally, both scenarios we used to test our hypothesis involve murder, which is an extreme act of violence. People might ascribe more freedom and responsibility in these extreme cases than in less emotional cases, so our results regarding bypassing and responsibility without freedom might not hold in less extreme scenarios. Further research should test whether similar patterns of results occur with less extreme wrongdoing as well as with responsibility or credit for good acts.

CRediT authorship contribution statement

Claire Simmons: Writing – original draft, Investigation, Conceptualization. **Paul Rehren:** Writing – review & editing, Formal analysis, Data curation, Conceptualization. **John-Dylan Haynes:** Writing – review & editing, Methodology, Supervision. **Walter Sinnott-Armstrong:** Writing – review & editing, Conceptualization, Funding acquisition, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Chalmers, D. J. (2011). Verbal disputes. *Philosophical Review, 120*(4), 515–566.

Clarke, R., and Capes, J. (2020). Incompatibilist (Nondeterministic) theories of free will. In Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition). URL = <https://plato.stanford.edu/archives/fall2020/entries/incompatibilism-theories/>.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). L. Erlbaum Associates.

Darrow, C. (1924). Plea for Leopold and Loeb. Voices of Democracy: The U.S. Oratory Project. https://voicesofdemocracy.umd.edu/clarence-darrow-plea-for-leopold-and-loeb-22-23-and-25-august-1924-speech-text/.

Dennett, D. (1983). Intentional systems in cognitive ethology: The "Panglossian paradigm" defended. *Behavioral and Brain Sciences, 6*(3), 343–355. https://doi.org/10.1017/S0140525X00016393

Dennett, D. (1984). *Elbow room*. MIT Press.

Feltz, A., Cokely, E. T., & Nadelhoffer, T. (2009). Natural compatibilism versus natural incompatibilism: Back to the drawing board. *Mind & Language, 24*(1), 1–23.

Figdor, C., & Phelan, M. (2015). Is free will necessary for moral responsibility? A case for rethinking their relationship and the design of experimental studies in moral psychology. *Mind & Language, 30*(5), 603–627.

Frankfurt, H. (1969). Alternate possibilities and moral responsibility. *The Journal of Philosophy, 66*(23), 829–839.

Gert, B., & Duggan, T. (1979). Free will as the ability to will. *Noûs, 13*(2), 197–217.

Ginet, C. (1997). Freedom, responsibility, and agency. *The Journal of Ethics, 1*(1), 85–98.

Hannikainen, I. R., Machery, E., Rose, D., Stich, S., Olivola, C. Y., Sousa, P., et al. (2019). For whom does determinism undermine moral responsibility? Surveying the conditions for free will across cultures. *Frontiers in Psychology, 10*(November), 2428.

Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2–3), 61–83.

JASP Team. (2020). *JASP (Version 0.16.1) [Computer software]*.

Jolliffe, I. T., & Jorge, C. (2016). Principal component analysis: A review and recent developments, 2015020220150202 *Philosophical Transactions of the Royal Society A, 374*. https://doi.org/10.1098/rsta.2015.0202.

Kenny, D. A., & Kenny, A (2003). *Action, Emotion and Will* (2nd ed.). Routledge. https://doi.org/10.4324/9780203711460

McCann, H. J. (2012). Making Decisions. *Philosophical Issues, 22*(1), 246–263.

McKenna, M., and Justin Coates, D. (2020). Compatibilism. In Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition). URL = <https://plato. stanford.edu/archives/win2020/entries/compatibilism/>.

Miller, J. S., & Feltz, A. (2011). Frankfurt and the folk: an experimental investigation of Frankfurt-style cases. *Conscious and Cognition, 20*(2), 401–414. https://doi. org/10.1016/j.concog.2010.10.015. Epub 2010 Dec 14. PMID: 2115952.

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*(July), 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*(2), 434–467.

Nahmias, E., & Murray, D. (2011). Experimental Philosophy on Free Will: An Error Theory for Incompatibilist Intuitions. In *New Waves in Philosophy of Action* (pp. 189–216). London: New Waves in Philosophy. Palgrave Macmillan.

Nichols, S., & Knobe, J. (2007). Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions. *Noûs, 41*(4), 663–685.

Pan, H., Liu, S., Miao, D., & Yuan, Y. (2018). Sample size determination for mediation analysis of longitudinal data. *BMC Medical Research Methodology, 18*(1), 32. https://doi.org/10.1186/s12874-018-0473-2

Pereboom, D. (2007). Harm Incompatibilism. In J. M. Fischer, R. Kane, D. Pereboom, & M. Vargas (Eds.), *Four Views on Free Will*. Oxford: Blackwell.

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

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

Sarkissian, H., Chatterjee, A., De Brigard, F., Knobe, J., Nichols, S., & Sirker, S. (2010). Is Belief in Free Will a Cultural Universal? *Mind & Language, 25*(3), 346–358. https://doi.org/10.1111/j.1468-0017.2010.01393.x

Sinnott-Armstrong, W. (2012). Free Contrastivism. In M. Blaauw (Ed.), *Contrastivism in Philosophy*. Routledge.

Stillman, T. F., Baumeister, R. F., & Mele, A. R. (2011). Free Will in Everyday Life: Autobiographical Accounts of Free and Unfree Actions. *Philosophical Psychology, 24* (3), 381–394.

Tingley, D., Teppei, H., Mit, Y., Keele, L., State, P., & Imai, K. (2014). Mediation: R Package for Causal Mediation Analysis. *Journal of Statistical Software., 59*. https:// doi.org/10.18637/jss.v059.i05

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

Vierkant, T., Deutschländer, R., Sinnott-Armstrong, W., & Haynes, J.-D. (2019). Responsibility without freedom? Folk judgments about deliberate actions. *Frontiers in Psychology, 1*–6. https://doi.org/10.3389/fpsyg.2019.01133

Vonasch, A. J., Baumeister, R. F., & Mele, A. R. (2018). Ordinary people think free will is a lack of constraint, not the presence of a soul. *Consciousness and Cognition, 60* (April), 133–151.

Vuorre, M., & Bolger, N. (2018). Within-subject mediation analysis for experimental data in cognitive psychology and neuroscience. *Behavior Research Methods, 50*(5), 2125–2143. https://doi.org/10.3758/s13428-017-0980-9

Wolf, S. (1990). *Freedom Within Reason*. New York: Oxford University Press.