

To appear in Oxford University Press volume on “Big questions on free will”

Free Will Without Metaphysics

Andrew E. Monroe and Bertram F. Malle

Brown University

Keywords: Free Will; Folk Concepts; Choice; Intentionality; Constraint; Blame; Moral
Judgment; Moral Responsibility

Abstract

Free will is the ability to transform thoughts and desires into actions, and this ability to act freely is presumed to underlie people's practice of moral praise and blame, as well society's practice of legal punishment. However, many scholars believe that people's concept of free will is hopelessly corrupted by metaphysical assumptions, such as a belief in the soul or a belief in magical causation. Scientific explanations of human behavior and the physical universe contradict such metaphysical assumptions. This raises the concern that science also invalidates the ordinary concept of free will and thereby unseats a key requisite for moral and legal responsibility. Such a concern, however, turns on the exact content of people's concept of free will. Here we discuss a program of research that seeks to clarify the folk concept of free will and its role in moral judgment. We also examine possible challenges to the concept and their impact on moral judgment. Our data show that people have a psychological concept of free will—they assume that “free actions” are based on choices that fulfill one's desires and are relatively free from internal and external constraints. Moreover, we argue that these components—choice, desires, and constraints—lie at the heart of people's moral judgments. Once these components are accounted for, the abstract concept of free will contributes very little to people's moral judgments, and challenges to the concept have little impact on those judgments.

Free Will Without Metaphysics

What does it mean to have free will? When asked, people widely believe that they have free will (Baumeister, Crescioni, & Alquist, 2010), and free will is commonly asserted as a critical underpinning for moral and legal responsibility (Greene & Cohen, 2004). But for such a seemingly widespread and important concept, there is remarkable confusion over its definition and use. Philosophers and theologians have debated the question of free will for millennia. Today, neuroscientists and psychologists have joined philosophers in trying to answer some nagging questions about free will: Is it an illusion (Wegner, 2002)? Is it incompatible with determinism (Nichols, 2011)? Can people be morally responsible without it (Greene & Cohen, 2004)?

However, what is the “it” in each of these questions? In all cases the “it” is the “folk” concept of free will. It is this concept that is suspected to be an illusion, incompatible with determinism, and required for moral responsibility. Unfortunately, scholars know very little about what constitutes this ordinary concept of free will. We therefore need clarity both on the concept and the underlying phenomenon, and in doing so we must go beyond philosophers’ and scientists’ intuitions. We must empirically examine ordinary people’s conceptualization of free will and their application of this concept in everyday life. Without taking seriously the actual folk concept of free will, any theory of free will is at “risk of having nothing more than a philosophical fiction as its subject matter” (Mele, 2001, p. 27). Science and philosophy might discover facts that suggest revisions to the folk concept of free will; but without knowing what the concept is we can hardly revise it.

In this chapter we present a program of research focused on developing an empirically grounded model of the folk concept of free will. Specifically, we address three questions

underlying the question of free will: What is people's concept of free will? How are free will and moral judgment related? And does threatening people's belief in free will affect social perception and moral judgment? Before delving into these questions, we offer a brief justification for the study of folk concepts.

People Believe All Kinds of Things

Some scholars dismiss the study of folk concepts. Many ordinary beliefs are culturally variable, inaccurate, and confused; why should we expect anything different for free will? But this view misunderstands the nature of folk concepts and their role in everyday life. Rather than being immature beliefs, folk concepts categorize phenomena and organize the relationships among categories (Malle, 2006). For example, people have a robust concept of intentionality (Malle & Knobe, 1997; Malle, Moses, & Baldwin, 2001) that is cross-culturally and cross-linguistically stable (Malle, 2008). It governs quick and effortless categorization of behavior (Barrett, Todd, Miller, & Blythe, 2005; Malle & Holbrook, 2012; Scholl & Tremoulet, 2000), explanations of behavior (Heider, 1958; Malle, 1999; Reeder, 2009; Woodward, 1998), and moral judgments (Darley & Shultz, 1990; Lagnado & Channon, 2008; Malle, Guglielmo, & Monroe, 2012; Ohtsubo, 2007; Young & Saxe, 2009, see also Dahourou & Mullet, 1999, for a replication with a non-Western sample). The intentionality concept is not so much a belief about facts in the world but, akin to a Kantian category, it fundamentally constitutes how people perceive the social world.

In the same way, people's folk concept of free will might structure how people perceive and respond to certain aspects of the social and moral world (see Baumeister, Masicampo, & DeWall, 2009; Vohs & Schooler, 2008). If we can empirically establish what this folk concept is

and how it functions in human cognition we can evaluate the troubling and widespread claims that it is imbued with metaphysics and a burdensome prerequisite for morality.¹

Common Claims About the People's Folk Concept of Free Will

Scholars of free will don't agree on many things, but they largely agree on what ordinary people's concept of free will is. Specifically, it is commonly taken to be a deeply metaphysical concept that involves magical thinking and rejects the normal laws of causality. Cashmore (2010) writes: "Free will makes 'logical sense,' as long as one has the luxury of the 'causal magic' of religion," but "neither religious beliefs, nor beliefs in free will, comply with the laws of the physical world" (p. 4502). Others echo this anti-scientific attitude: "Free will is the idea that we make choices and have thoughts independent of anything remotely resembling a physical process" (Montague, 2008, p. R584). "The jargon of free will in everyday language... requires us to accept local pockets of indeterminism in an otherwise deterministically conceived world view" (Maasen, Prinz, & Roth, 2003, p. 8).

If these characterizations of the folk concept are correct, then what people subscribe to runs counter to science and is apt to be overturned. But overturning free will presents a problem. Free will is commonly viewed as necessary for moral and legal responsibility. For example, Greene and Cohen (2004) claim that the law is predicated on a libertarian assumption of free will. Darwin makes a bolder claim, arguing that without a belief in free will "one deserves no credit for anything... nor ought one to blame others" (Darwin, 1840, p. 27). Therefore, if science undermines the existence of free will, then the justification for our moral and legal practices is forfeit. This threat, however, rests on the critical assumption that people have a metaphysical concept of free will. Yet, before we charge people with holding such a confused concept, and before we declare their mental and moral practices corrupt, we need scientific evidence.

An Empirical Investigation of Free Will

Investigations of free will have recently garnered widespread popular and scientific attention. These studies, however, often focus on some variation of the Libet experiments (Filevich, Kühn, & Haggard, 2013; Haggard, 2011; Lau, Rogers, Haggard, & Passingham, 2004; Libet, Gleason, Wright, & Pearl, 1983; Schurger, Sitt, & Dehaene, 2012; Trevena & Miller, 2010) or on probing people's intuitions regarding whether free will and moral responsibility are compatible with determinism (e.g., Nahmias, Morris, Nadelhoffer, & Turner, 2005; Nichols & Knobe, 2007). Empirical investigations into people's conceptualization of free will are virtually non-existent.

In a first study on the topic, Monroe and Malle (2010) probed people's concept of free will by inviting them to report "what you think it means to have free will." This approach mirrored Malle and Knobe's (1997) research, which successfully demonstrated that the criteria for a concept could be elicited by asking people to explicate the concept (e.g., "When you say that somebody performed an action intentionally, what does this mean?" p. 106).

Monroe and Malle's (2010) findings strikingly diverged from the widespread claims about the folk concept of free will. Metaphysical commitments were all but absent in the data. Out of the 180 participants tested, and the 259 total responses collected, only a single participant gave a response that fit a metaphysical interpretation of free will: "Free will is when you can make a decision that is completely untouched by outside factors" (Monroe & Malle, 2010, p. 216). In the remaining 258 statements, people converged on a psychological definition of free will. They defined free will as (a) the ability to make a choice; (b) acting consistent with one's desires; and (c) being (reasonably) free of constraints (see Table 1). Importantly, the constraints that people mentioned were psychological in nature and referred to such factors as peer pressure

and social status (e.g., “[making] decisions without fear and overriding influence from others”; “To be able to say and do whatever you want no matter your race, IQ, or finance [sic] situation”).

Table 1.

Monroe & Malle's (2010) content coding of folk definitions of free will

Coding Categories	Percent of Participants Mentioning Each Category
Ability to make a decision or choice	65%
Doing what you want	33%
Acting without constraints	29%

The goals of this first study were modest. We set out to empirically document people’s conceptualization of free will, and it appears that people hold a psychological, not a metaphysical concept of free will. Though a single study is insufficient to bolster this conclusion, additional data have recently emerged that are consistent with Monroe and Malle's (2010) findings.

In one such study, Stillman, Baumeister, and Mele (2011) asked participants to produce an autobiographical account of actions they felt were either performed “of their own free will” or “not the result of free will.” The results paralleled those of Monroe and Malle (2010). People in the “free will” condition reported behaviors associated with pursuing goals, making choices, and acting against external forces (e.g., temptation or pressure from others). By contrast, participants in the “no free will” condition wrote about behaviors under constraint, such as in the presence of powerful authority figures. Metaphysical commitments were tellingly absent in both conditions.

These preliminary data cast doubt on the characterization of people’s concept of free will as magical and metaphysical. Instead they suggest that the ordinary understanding of free will is rooted in the folk concept of intentionality (especially the components of desire and choice) and extends beyond it by also considering internal and external constraints on behavior. However, the studies by Monroe and Malle (2010) and Stillman et al. (2011) share two limitations. First,

both rely on undergraduate student participants and therefore may not be representative of the population at large (Henrich, Heine, & Norenzayan, 2010). Second, both rely on people's *explicit* reports on the concept of free will. Thus, one could argue that people's concept is still *implicitly* metaphysical; people simply fail to report these implicit metaphysical commitments. Below we bring empirical data to bear on each of these claims. We first consider the possibility that the concept of free will espoused by university students fails to replicate in a general population sample.

Free Will in the Community

In a new study, we explored people's folk concept of free will by employing a structured interview of community members. We randomly selected individuals from the phone book who lived within a 5 mile-radius of campus and invited them into the lab for a paid study. These 39 participants (19 female) were older ($M = 34.2$, $SD = 15.5$) than a typical undergraduate sample, and though they self-identified as considerably liberal ($M = 2.4$ $SD = 1.12$, on a 1-7 scale), they were evenly split with regard to identifying themselves as religious (20 did, 19 did not). The interviews were audio recorded and later transcribed by a research assistant who was blind to the study's goals. Three independent coders classified participants' open-ended responses to the interview questions.

In addition to putting Monroe and Malle's (2010) finding to the test with a more representative sample, we also wanted to probe several additional aspects of people's concept of free will, including its connection with other potentially related concepts (e.g., free choice, acting freely, intentionality), its application (e.g., building a novel agent that has free will), and specific beliefs about free will (e.g., whether it is inborn or develops over time).

The concept. If people’s concept of free will is indeed a psychological concept featuring choice, desire, and lack of constraints, then community members should define free will similarly to the college students in Monroe and Malle (2010). Indeed, we replicated this conceptual structure (Table 2), again without mention of souls, indeterminism, or other metaphysical commitments.

Table 2.

Components of community participants’ lay definitions of free will

Coding categories	Percent of Participants
Choice	41%
Desires	38%
Forethought	26%
Free from (external or internal) constraints	74%
Reference to metaphysics, souls, or indeterminism	0%

In two respects, community participants’ lay definitions differed from those of university students in Monroe and Malle (2010). First, community participants cited the absence of constraints much more frequently (74%) than student participants (29%). This perhaps speaks to a difference in the salience of constraints that emerges with age. While college students experience relatively few constraints on their behavior, older community members may be keenly aware of the various constraints impinging on their lives (e.g., bills, jobs, children). Second, community members mentioned an additional component of free will that was absent in the student sample—forethought. This component was exemplified by “weighing the benefits of action,” “premeditation,” or “having thought out one’s actions.” This criterion presupposes the process of choice but serves to characterize the type of choice as being thoughtful or made in consideration of the future. Such a component is consistent with our contention that people’s

conception of free will is an extension of the folk concept of intentionality, in which considerations of forethought and awareness play a considerable role (Malle & Knobe, 1997).

Relations to other concepts. In addition to defining free will, participants were asked to define several other concepts that were hypothesized to relate to free will: free choice, acting freely, and acting intentionally. With regard to free choice, the majority of participants (56%) explicitly linked free will and free choice. Whereas free will was described as a general capacity, free choice was defined as a concrete demonstration of one's free will (e.g., "Free choice is almost like a slight step down from free will." "Free choice is sort of the same as free will where they made a decision based on what they knew and how they felt."). In addition to the explicit link between the two concepts, people defined free choice as having similar qualities to free will including being free from constraints (56%), making a choice (21%), acting on personal desires (20%), and forethought (13%). A unique component of free choice was that of options. Participants defined free choice as requiring behavioral alternatives (e.g., "they had several options available to them").

Participants' definitions of "acting freely" largely focused on the lack of constraints (56%) and making choices (56%). A subset of participants (13%) interpreted acting freely as spontaneity—not in a metaphysical sense, but as acting impulsively (e.g., "not thinking through carefully"; "acting recklessly"). Definitions of acting intentionally mirrored Malle and Knobe's (1997) conceptual structure, with participants reporting that acting intentionally required a desire for a particular outcome (26%), a belief about how to bring about said outcome (35%), deliberating and intending to act (31%), and to a lesser extent, being aware of one's actions (9%).

Applications. Going beyond semantic intuitions, we also asked community participants to indicate how they would “build” an agent that has free will (“If you wanted to build a biological organism [or a robot] that had free will, what abilities would it need to have?”). Choice was again the dominant category, with 81% of participants citing it as a necessary ingredient for an agent to have free will (e.g., “for it to be naturally free will it would have to be able to choose”). The category with the second-highest prevalence (35%) was a capacity for autonomy – defined as being able to resist constraints (“It would have to be able to choose to not be persuaded by an external factor”). Just under a quarter of participants mentioned either consciousness (22%) or being ambulatory (22%). However, these capacities were usually mentioned in service of choice (e.g., “. . .but also to have some sort of consciousness in which it could actually think and consciously reason to come to the decisions”) or for carrying out chosen actions (e.g., “It would need to be able to move around relatively freely or have a chance to—for me—that it could have a chance of carrying out its will”). To a lesser extent people mentioned needing desires, goals, and preferences (16%) or moral principles (14%) in order to have free will, but once more, they failed to mention any conditions that could count as metaphysical requirements (e.g., a soul, uncaused causer).

Specific beliefs about free will. We also asked people to express some beliefs they had about free will—revealing more their *conception*, not their *concept* of free will. These responses provide further evidence for the claim that free will in ordinary people’s mind is a psychological process. Asked whether free will is something that humans are born with or develops with age, the majority of participants (71%) reported that an agent’s capacity for free will develops over the lifespan, compared with 21% who viewed free will as an innate, unchanging module. For example, one participant wrote: “I think it develops with age. You’re born with some free will

but it's more just biological actions, not something you think about... But I think as you develop, you develop opinions and the ability to think for yourself.” Conversely, when asked whether something could “take away the capacity for free will,” 94% of participants answered yes. When asked specifically what factors could take away free will, people reported coercion (63%), brain damage (40%), and physical limitations (37%), such as paralysis.

These results, in conjunction with the previous research by Monroe and Malle (2010) and Stillman et al. (2011), show considerable support for the claim that the folk concept of free will is fundamentally psychological, not metaphysical. Community members and college students alike identify choice and a lack of constraint as the core components of free will. Moreover, this pattern emerges when people define the concept (e.g., what does it mean to have free will?) and when they apply the concept (e.g., what capacities would an agent need to have free will?).

Yet these findings will not satisfy the skeptic who might insist that people have *implicit* commitments to metaphysical properties of free will (e.g., being an uncaused causer). Unfortunately, no study that fails to provide evidence for metaphysical commitments can satisfy the skeptic; for there might always be some better, more clever way to expose those cloaked commitments. In this way, the claim of metaphysical commitments is dangerously close to an unfalsifiable hypothesis. Nonetheless, as dogged empiricists we must keep trying to put the hypothesis to a test. In so doing, two challenges await. First, most assessments of implicit cognition rely on language (e.g., techniques of unscrambling or semantic priming), but the candidate metaphysical assumptions about free will are difficult to formulate in ordinary language: How do we translate charges such as “uncaused causer”? “contra-causal will”? “nondeterminism”? Second, no clear criteria have been offered for identifying something as

“metaphysical.” Is *counterfactual* reasoning metaphysical? Is a choice from among *options* metaphysical?

In two recent studies we tried to at least partially address these challenges. First, we examined two properties that most would consider metaphysical: breaking the causal flow of the universe and being an uncaused causer. Second, rather than gathering explicit definitions of each property we asked people to make judgments about whether various behaviors instantiated the property, and we assessed the speed with which they made those judgments. The guiding assumption was that people are facile at making judgments of such properties as choice and intentionality (Malle & Holbrook, 2012); if they (implicitly) consider certain agents as “uncaused causers” or consider behaviors as “breaking the causal flow of the universe” then they should also be fast and facile at making those judgments.

We selected these particular contentions because they are apt exemplars of the metaphysical portrayal of free will that is extremely common in the literature. For example, according to Bayer, Ferguson, and Gollwitzer (2003), people believe that “willfulness somehow springs forth from some special uncaused place” (p. 100).

To test these hypotheses we adopted a paradigm used by Malle and Holbrook (2012) to investigate simultaneous inferences. Participants listened to short descriptions of intentional behaviors (e.g., “Kaylee took money from her mom's wallet”) and unintentional immoral behaviors (e.g., “Yolanda broke her grandmother's heirloom vase”), and they pressed a *yes* or *no* key in response to one of several inference cues. Both of our studies included the cues: INTENTIONAL? FREE WILL? CHOOSE? BREAK? OPTIONS?. In our second study we added the cue UNCAUSED? to test an additional metaphysical exemplar. Participants were trained on the meaning of each of the cues. For example, participants learned that whenever they

see the cue “BREAK” it means “did the person *break* the causal flow of the universe?” The overall likelihood of responding with *yes* to each inference cue and the average time to respond with *yes* to each inference cue were the critical dependent variables.²

The data showed a clear pattern. Replicating previous results (Malle & Holbrook, 2012), people were fast at making inferences of intentionality, $M = 1301$ ms ($SD = 367$). Moreover, inferences of choice ($M = 1223$ $SD = 406$) and having options ($M = 1249$ $SD = 360$) were at least as fast, if not faster than intentionality. The key question in this study was whether inferences of free will would cluster with inferences of intentionality, choice, and having options (suggesting that people have a psychological concept of free will) or with the property of BREAK (suggesting that people have an implicit metaphysical commitment to indeterminism). The data back the hypothesis of a psychological concept of free will. People’s speed to infer free will ($M = 1244$, $SD = 386$) clustered with intentionality, choice, and options, whereas judgments of causal breaks ($M = 1550$ $SD = 375$) were significantly slower, $p < .01$. Our second study replicated this pattern of findings, and it also showed that, compared to inferring free will, intentionality, choice, or options, people were significantly slower ($p < .01$) to infer that an agent was an “uncaused cause” of her actions (see Figure 1).

Examining the likelihood of participants responding with *yes* to a given cue showed stark differences in the rates people affirmed the different cues (See Figure 2). In both studies people were near ceiling in their affirmation of the folk psychological cues (FREE WILL? INTENTIONAL? CHOOSE? OPTIONS?), while affirmation of the metaphysical cues (BREAK? UNCAUSED?) was significantly lower. Together with the reaction time results, this shows that while people are able to answer the metaphysical questions when pressed, those concepts are far from intuitive and also not closely related to people’s concept of free will.

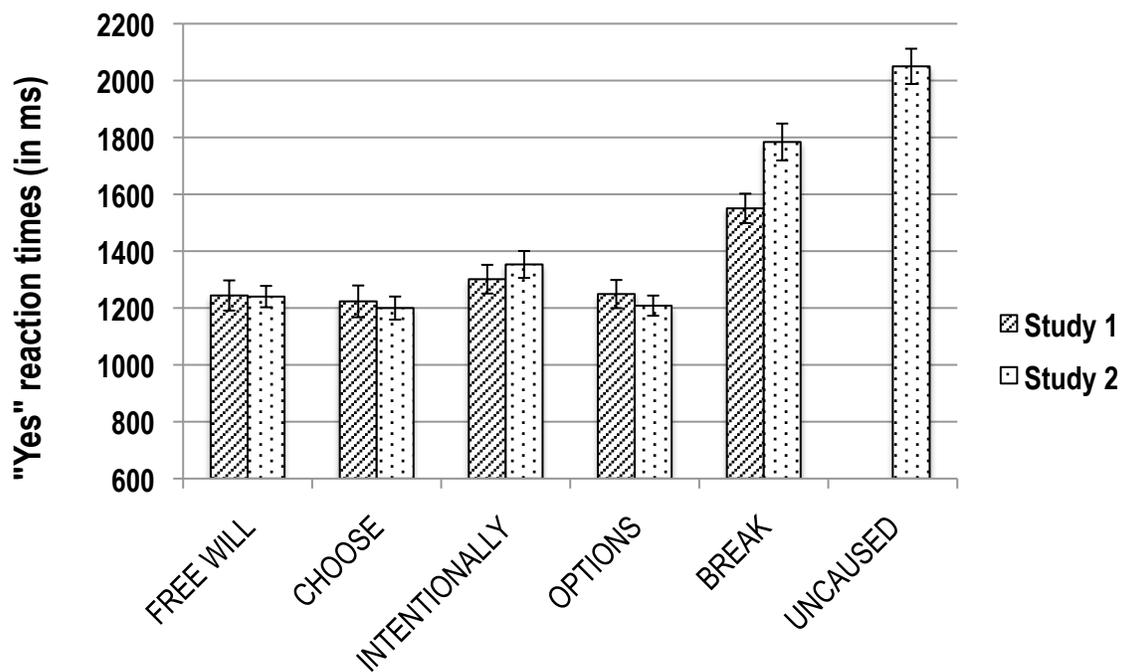


Figure 1. Reaction times for "yes" responses to cues.

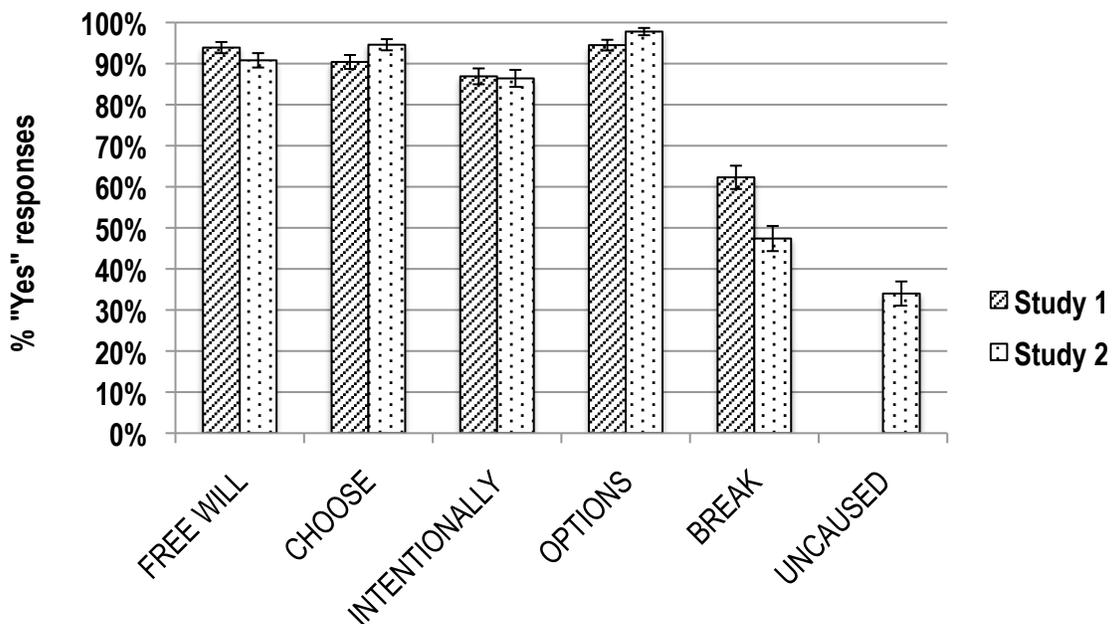


Figure 2. Likelihood of "yes" responses to cues.

These first empirical investigations of the folk concept of free will have revealed a number of important findings. First, we repeatedly demonstrated that people's concept of free will lacks the strong metaphysical commitments commonly attributed to it. Rather, these data support the claim that people hold a psychological concept of free will. Both university students and community members explicitly characterize free will as choice, acting intentionally, considering options, and acting free of (or overcoming) constraints. Moreover, these explicit reports are confirmed by several reaction time studies. Judgments about free will, choice, intentionality, and options form a tight conceptual bundle, while metaphysical concepts such as breaking deterministic laws and being an uncaused causer are outside of people's everyday concept.

Is Free Will Needed for Moral Judgment?

Thus far it seems that the best way to characterize people's concept of free will is as a concept without metaphysics. However, maybe we have been asking the wrong questions. Perhaps people's metaphysical commitments surface when we probe the connection between free will and moral judgments.

Free will is, in part, viewed as a Big Question because it is assumed to undergird everyday morality. This assumption is typically interpreted to imply that if an agent did not act of her own free will, then it is inappropriate to blame or punish her. "The concept of free will most philosophers are interested in is the one that is necessary for moral responsibility and attributions of praise and blame" (Nahmias et al., 2005, p. 576). Is this also true for people's folk concept of free will?

The connection between free will and morality appears to be fertile ground for study. Previous research examining free will beliefs and moral judgment has shown that metaphysical

considerations, such as highlighting the presence of a deterministic universe, can produce variability in people's blame judgments (See Nahmias, 2006; Nahmias, et al., 2005; Nichols, 2006; Nichols & Knobe, 2007; Roskies & Nichols, 2008). Thus, by examining the connection between free will and moral judgments we can perhaps unearth metaphysical commitments lurking in people's free will concept.

One commitment commonly attributed to people is that free will requires the presence of a soul as a "first mover" or "magical cause." Bargh and Earp (2009) claim that people's concept of free will is "laden with the concept of a soul, a non-physical, unfettered, internal source of choice-making" (p. 13). However, Nahmias, Coates, and Kvaran (2007) showed that very few people (15-25%) agree with the statement "Humans have free will only because they have nonphysical souls." In our interview study we also found people reluctant to claim that a soul is necessary for having free will. When asked explicitly about such a necessity, about one third of participants (28%) affirmed that a soul is necessary for free will; a third (36%) denied any relationship between having a soul and free will; and the final third (36%) were unsure whether the soul was needed for free will. But whatever people reported explicitly, it is still possible that, for many people, a belief in a soul reveals an unconscious metaphysical commitment that guides ascriptions of free will and, therefore, their moral judgments.

In a recent series of studies we set out to explicitly test the claim that having a soul is necessary for free will (Bargh & Earp, 2009; Cashmore, 2010). If correct, then people's willingness to ascribe free will to an agent should depend on whether that agent is believed to have a soul. Additionally, insofar as free will is necessary for moral responsibility, ascriptions of a soul should also be predictive of people's judgments of blame. We contrasted this set of predictions with Monroe and Malle's (2010) proposal that people have a psychological concept

of free will. On this view, while many people may indeed believe in souls, such beliefs are irrelevant for judgments of free will and blame. Rather, ascriptions of free will should depend primarily on an agent's perceived capacity for choice and intentional action, and these capacities should also predict blame judgments.

In ordinary human agents, however, ascriptions of souls, choice, and free will are typically confounded. To disentangle these properties, and to test the relationship between ascriptions of choice and a soul to free will and morality, we conducted two studies. In the first study we constructed five different agents whose descriptions varied in a number of features, most notably in either having or lacking a human brain, a human physiology, and the capacity to make choices. Participants read one of the five agent descriptions, made blame judgments about various norm-violating actions the agent performed (e.g., throwing a water balloon off a theatre balcony), and were invited to judge whether the agent had various capacities, including a soul, choice, and free will.

The results showed that people based their decision to grant an agent free will on the agent's perceived capacities to make choices and to act intentionally, not on the possession of a soul. For example, people granted souls only to human agents—both a normal human and a human suffering from “Crick's disorder,” which eliminated the person's capacity to make choices. Of these two human agents, however, only the normal human was granted free will. Conversely, a cyborg (a human brain in a robot body) was denied a soul but granted the ability to make choices, and people therefore ascribed free will to this agent. Similarly, the only capacities that mattered for judging an agent blameworthy were the capacity for intentional action, choice, and lack of external constraints; having or lacking a soul did not predict judgments of blame.

We replicated these results in a second study in which participants read one of four agent descriptions that were pretested to explicitly manipulate (a) having the capacity for choice and (b) having a soul: Normal human (choice present/soul present), Cyborg (choice present/soul absent), Akkratic Human (choice absent/soul present), or Robot (choice absent/soul absent). The data confirmed the results of the previous study. Ascriptions of free will were largely predicted by intentionality and choice (62% explained variance) rather than having a soul (8.5% explained variance), and intentionality and choice explained the majority of the variance in blame judgments (54%), while soul ascriptions failed to explain any variance at all.

The two studies also revealed that free will might not be as important for moral judgment as previously thought. Free will only weakly predicted blame judgments, and its predictive power was further reduced when controlling for the predictive power of intentionality and choice. In the first study, free will did not account for any unique variance beyond intentionality and choice, and in the second study it explained only 2% unique variance in blame judgments.

These studies suggest two conclusions that are consistent with Monroe and Malle's (2010) psychological account of free will. First, the perceived presence of a soul is neither necessary nor sufficient for people to ascribe free will. It is not necessary because a cyborg without a soul was granted free will; and it is not sufficient because a human with a disorder that disrupts choice is granted a soul but not free will. By contrast, the perceived capacities for choice and for intentional action are jointly necessary conditions for ascribing free will.³ Second, free will is, by itself, not necessary for blame. Only insofar as “free will” is a shorthand for ascribing intentional agency, choice, and lack of constraints does it predict blame. Once ascriptions of choice and intentional action are taken into account, free will ascriptions contribute nothing new to blame judgments.

Once again, however, we must confront the specter of the implicit; people might still have some sort of implicit metaphysical beliefs about free will. One way to elicit those beliefs is to challenge them and examine the effect on observable behavior (Baumeister et al., 2009; Nichols & Knobe, 2007; Rigoni, Kühn, Gaudino, Sartori, & Brass, 2012; Stillman & Baumeister, 2010; Vohs & Schooler, 2008). If challenging people's (presumed) commitment to indeterminism, the soul, or magical causation alters their moral judgments, then that would be evidence for metaphysical commitments in people's moral judgments.

Can Threatening Free Will Alter Social Perception and Moral Judgment?

To test this hypothesis we presented participants at a local public beach ($N = 275$) with one of six challenges to free will or one of two control conditions: a pro-free will statement or no statement. The majority of the free will challenges were experimental manipulations from previous research (Baumeister et al., 2009; Nahmias et al., 2005; Nichols & Knobe, 2007; Vohs & Schooler, 2008); one was an implied threat to free will (Greene & Cohen, 2004), and we derived one unique challenge from responses in our interviews.

Participants were presented with one of the seven statements and then read about an agent who committed a moral violation.⁴ For example, in one condition participants were presented with the following text:

People are nothing but a pack of neurons. Their joys and sorrows, memories and ambitions, and their sense of self are no more than the activation of a vast assembly of nerve cells. All of people's decisions and actions are completely caused by a particular pattern of neurons firing in the brain – demonstrating that such things as beliefs, desires, and values don't in fact guide decisions.

One afternoon Sean stole expensive clothing from a department store. As always, the behavior was completely caused by a particular pattern of neurons firing.

After reading the statement and the moral violation, participants rated the blameworthiness of the agent and the wrongness of the action. In addition, participants were

asked to decide which of eight characterizations accurately described the agent’s behavior. Four were phrased in folk-psychological language (e.g., “acted intentionally”; “had free will”; “made a choice”; “acted freely”), and four characterizations were phrased in what may be considered metaphysical language (e.g., “made a ‘break’ in the causal laws of the universe”; “was fully determined”; “couldn’t have acted differently”; “the decision was caused by his brain”).

Participants were asked to evaluate each characterization of the agent’s behavior and decide if it sounded “certainly correct,” “maybe correct,” or “not at all correct.” Finally, participants indicated whether they had found the initial statement (one of the seven challenges or pro-free will statement) persuasive.

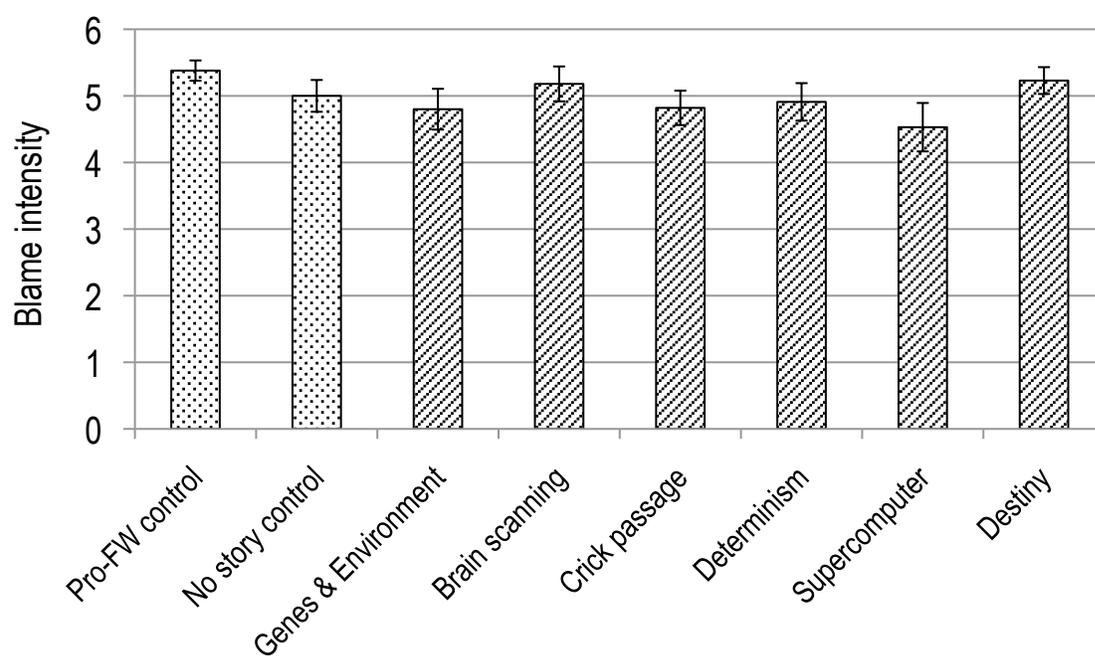


Figure 3. *Blame ratings across free will threat and control conditions*

We found no effect of condition on judgments of blame and wrongness (see Figure 3 for blame results). That is, no claim about the nonexistence (or existence) of free will altered people’s moral judgments, compared with the control condition. Additionally, in spite of the

free will challenges, people strongly endorsed the folk-psychological characterizations (i.e., choice, acting freely, and intentionality) of the agent's action. On average, 75% of participants considered the folk-psychological characterizations "certainly correct." By contrast, only 12% of participants considered the metaphysical characterizations "certainly correct." Neither the lack of effects on moral judgments nor the preference for folk-psychological property ascriptions can be explained by participants' rejection of the challenges as unpersuasive. The rated persuasiveness of the challenges was unrelated to people's moral judgments and unrelated to their endorsement of the characterizations of the agent's behavior ($ps > .4$).

Thus, people strongly hold that agents act intentionally, make choices, have free will, and are morally responsible, even in the face of challenges to beliefs in free will. This finding might suggest that people's social and moral judgments are recalcitrant to just about any information. But that is clearly not the case. In one study, for example, we manipulated whether an agent met *eligibility* criteria (*cf.* Roskies & Malle, 2013) for being morally responsible, such as maturity, understanding the wrongness of one's actions, and the ability to make (or inhibit) choices. We found dramatic reductions in blame, particularly when an agent was unable to make choices or understand the wrongness of his actions.

In a second study (Monroe, 2013) we manipulated an agent's capacity for choice and examined people's perceptions of intentionality and judgments of blame of the agents. Participants were presented with descriptions of different agents (one at a time) who seriously harmed a stranger. The agent descriptions manipulated (within subjects) whether the agent's choice capacity was disabled entirely (e.g., because of psychosis, brain abnormalities, hypnosis), disabled only partially (e.g. because of abuse, social pressure, powerful emotions), or it was intact but harm was caused in an unconventional manner (e.g., brief seizure). The results

showed that people blamed the agent strongly when choice capacity was intact but harm was unconventionally caused. Blame was reduced when choice was partially disabled, and even more so when it was entirely disabled (See Figure 4). Further, blame within each of the conditions was strongly predicted by the perceived intentionality of the agent's action ($ps < .01$).

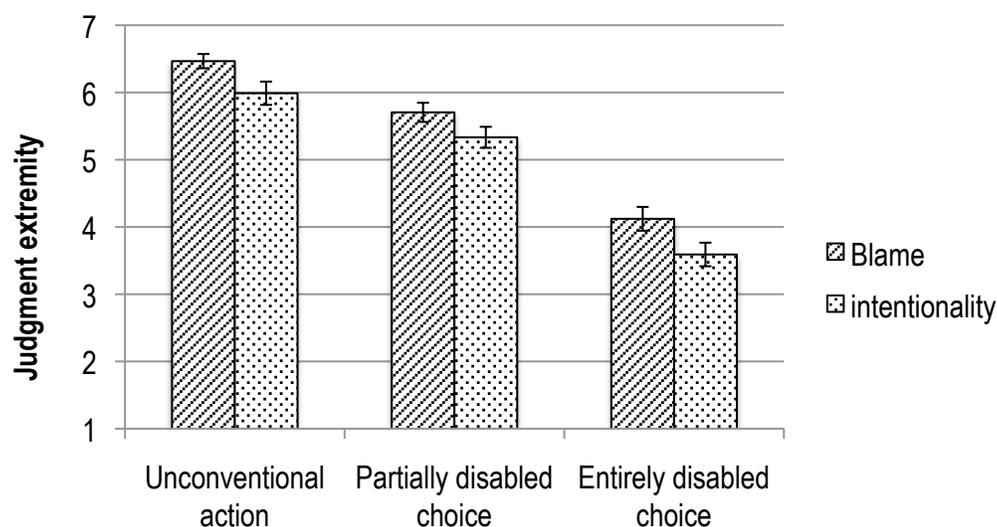


Figure 4. Blame decreases in response to limitations on choice

Thus, we may conclude that the earlier reported “free will threats” challenged aspects of free will that are irrelevant for social perception and moral judgment. By contrast, once relevant features of mind and behavior are challenged—such as choice and intentionality—blame is altered predictably.

Unanswered Questions

Several questions remain unanswered by our studies. For one thing, we may have failed to convince the skeptic who suspects lurking metaphysical commitments underneath the folk concept of free will. If there are such commitments, however, they must be empirically detectable. We therefore can only urge researchers to develop new methods to probe the concept of free will for implicit, unconscious components that our studies have not detected. The criteria

are clear: there must be some evidence for a commitment; it must be arguably metaphysical (i.e., contradicting laws of nature); and it must be predictive of or necessary for ascriptions of free will, choice, or moral judgment. In our own assessment, a soul, uncaused causes, or indeterminism are unlikely candidates of such commitments. However, there is at least one conceptual component that we have not yet examined in detail—the principle of alternative possibilities (as philosophers call it) or the ordinary assumption that sometimes agents could have acted differently. From our data it appears that something like the availability of “options” is closely related to free choice (see interview study) and to free will (see reaction time studies). But what aspect of “having options” is important to people? Is it that the agent could have acted differently even if everything up to the moment of action was identical except that the agent “chose” to go in a different direction? This seems rather unlikely. Given that people make sense of decisions and actions by looking for an agent’s *reasons* that generated the decision to act (Malle, 1999), if the agent chose A over B, then his reasons-for-A were involved in generating the choice whereas if he chose B then his reasons-for-B were involved in causing the choice. Therefore, not everything can be identical in two worlds in which an agent chooses two different actions; the agent’s reasons that (at least partially) generated the choice, must be different as well.

Alternatively, people may demand that an action is free (and subject to moral evaluation) only when there was an actual possibility that the person could have acted otherwise—that is, there was at least one possible world with the same past and the same laws of nature in which the agent did not do what she actually did. If there *is not* such a possibility, then the action wasn’t free (Frankfurt, 1969). To test this hypothesis Miller and Feltz (2011) presented participants with a number of scenarios in which an agent committed a car theft even though he could not, in

reality, have done otherwise. In Study 1, for example, Mr. Jones had a neural implant programmed to cause his decision to steal a car at a specific time, just in case Jones did not decide, on his own, to steal the car. As it so happens, he did decide on his own. Was he morally responsible for deciding to steal the car? On a Likert scale ranging from 1 (not at all agree) to 7 (strongly agree), participants' mean was 5.6, and about 2/3 of them were on the "agree" side of the scale (above the midpoint). So people don't seem to demand that the agent could have (in an otherwise parallel reality) acted differently; what matters is that in *this* reality the agent made a choice and that this choice caused the action.

It would be interesting to let the alternative scenario play out—where the neural implant "causes the agent's decision" to steal the car. Because an agent's normal decision-making process (which includes forming an intention on the basis of reasons) is, on our model, essential to free action, people's blame should drop considerably under this condition. "Having options" or "alternative possibilities" may thus be no more than a psychological description of an agent's situation before choosing to act: multiple subjectively possible and reasonable paths of action and the capacity to use "normal" decision making faculties in selecting one of the paths. Even if in reality all paths but one are blocked (but the agent doesn't know that), actions that are based on normally unfolding decisions from among such options are considered free. If the subjectively represented options are limited—as in the prototypical case of a gun to one's head or strong pressure from authority—a decision is made, but the person "could not have" reasonably decided otherwise and therefore did not freely act.

This analysis can be pushed further toward cases in which an agent in fact did not intentionally bring about an outcome (e.g., a driver accidentally injuring a pedestrian). Here people blame the agent if they conclude that he *could have* prevented the outcome (Malle et al.,

2012). How do they arrive at this conclusion? This is an empirical question that has received far too little attention, in part because the exact process of constructing counterfactuals is not well understood. Perhaps people simulate the decision situation—and if, in this simulation, clear and reasonable options emerge that the agent must have rejected (e.g., deciding to not try pulling the accident victim out of the car), people consider the agent blameworthy. But are people committed to a parallel possible world? To a nondeterministic interpretation of the universe? We think not. But we don't yet have the data to be confident.

The Myth and Reality of Free Will

We began this chapter with a question: What does it mean to have free will? We presented the initial results of a research program aimed at answering this question. Though the research is surely incomplete, there are several conclusions we are inclined to draw about how people conceptualize free will and how they use this concept in everyday life.

What is people's concept of free will? The data presented here support the view that people have a psychological concept of free will—essentially, free will is choice and intentional action, without constraint. Moreover, each study we conducted failed to provide evidence for the widespread claim that people are committed to a metaphysical notion of free will. A skeptic might insist that we have not asked the right questions or that people's commitments are deeply intuitive and implicit. But by offering empirical evidence where before there has only been scholarly conjecture, we argue that the burden of proof is shifting toward those who claim that people have a metaphysical concept of free will.

What is the relationship between free will and moral judgment? Moral judgments are widely thought to presuppose free will—people do not blame (or praise) an agent who lacks free will. In one sense this is correct. The capacity for free will is necessary for ascribing moral

blame insofar as it is a summary label for the capacities for choice and intentional action. But our data suggest that once these two constituents of free will are accounted for, there is no further role for a unique free will capacity to play. There is one way in which “free will” goes beyond choice and intentionality: as freedom from constraints. This, the third component of the psychological folk concept of free will (Monroe & Malle, 2010), makes a unique contribution. Even when an agent made a choice and acted intentionally on it, if the agent had no reasonable alternative option, the agent is not to blame.

What are the social implications of threatening free will? Threats to the belief in free will cause people to behave in socially maladaptive ways (e.g., heightened cheating and aggression, less helping, reduced learning; Baumeister et al., 2009; Stillman & Baumeister, 2010; Vohs & Schooler, 2008). So far, no parallel effects have been found in the domain of social perception: threats to the belief in free will do not seem to change people’s judgments about intentionality, choice, freedom, or morality. More research is needed to reconcile these results (Schooler, Vohs, Nahmias, & Nadelhoffer, 2013). It is possible, for example, that effects on behavior are mediated by moral disengagement, diminished self-control, or ego depletion, whereas moral judgments may not be susceptible to these processes in the same way. Moral judgments may simply involve consideration of choice, intentionality, and preventability, and none of these require a special belief in free will. It may even turn out that a belief in free will guides action (Baumeister et al., 2010), whereas the folk-psychological concepts of choice and intentionality guide social and moral judgment. But one thing is for certain: we need empirical data to resolve these questions; discussions about whether free will “exists,” or scholarly intuitions about the folk concept, will not suffice.

References

- Bargh, J. A., & Earp, B. D. (2009). The will is caused, not “free.” *Dialogue, Newsletter of the Society for Personality and Social Psychology*, 24, 13–15.
- Barrett, H. C., Todd, P. M., Miller, G. F., & Blythe, P. W. (2005). Accurate judgments of intention from motion cues alone: A cross-cultural study. *Evolution and Human Behavior*, 26, 313–331. doi:10.1016/j.evolhumbehav.2004.08.015
- Baumeister, R. F., Crescioni, A. W., & Alquist, J. L. (2010). Free will as advanced action control for human social life and culture. *Neuroethics*. doi:10.1007/s12152-010-9058-4
- Baumeister, R. F., Masicampo, E. J., & DeWall, C. N. (2009). Prosocial benefits of feeling free: Disbelief in free will increases aggression and reduces helpfulness. *Personality and Social Psychology Bulletin*, 35, 260–268. doi:10.1177/0146167208327217
- Bayer, U. C., Ferguson, M. J., & Gollwitzer, P. M. (2003). Voluntary action from the perspective of social personality psychology. In S. Maasen, W. Prinz, & G. Roth (Eds.), *Voluntary action; brains, minds, and sociality* (pp. 86–107). New York, NY US: Oxford University Press.
- Cashmore, A. R. (2010). The Lucretian swerve: The biological basis of human behavior and the criminal justice system. *Proceedings of the National Academy of Sciences*, 107, 4499–4504. doi:10.1073/pnas.0915161107
- Dahourou, D., & Mullet, E. (1999). The relationships among intent, consequences, and blame in Burkina Faso adolescents and young adults. *IFE Psychologia: An International Journal*, 7, 32–45.

- Darley, J. M., & Shultz, T. R. (1990). Moral rules: Their content and acquisition. *Annual Review of Psychology*, *41*, 525–556. doi:10.1146/annurev.ps.41.020190.002521
- Darwin, C. R. (1840). Old and useless notes about the moral sense and some metaphysical points. (P. H. Barrett, Trans.).
- Filevich, E., Kühn, S., & Haggard, P. (2013). There is no free won't: Antecedent brain activity predicts decisions to inhibit. *PLoS ONE*, *8*, e53053. doi:10.1371/journal.pone.0053053
- Frankfurt, H. G. (1969). Alternate possibilities and moral responsibility. *Journal of Philosophy*, *66*.
- Greene, J. D., & Cohen, J. D. (2004). For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*, 1775–1785. doi:10.1098/rstb.2004.1546
- Haggard, P. (2011). Decision time for free will. *Neuron*, *69*, 404–406. doi:10.1016/j.neuron.2011.01.028
- Heider, F. (1958). *The psychology of interpersonal relations*. New York, NY US: Wiley.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83. doi:10.1017/S0140525X0999152X
- Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, *108*, 754–770. doi:10.1016/j.cognition.2008.06.009
- Lau, H. C., Rogers, R. D., Haggard, P., & Passingham, R. E. (2004). Attention to intention. *Science*, *303*, 1208–1210. doi:10.1126/science.1090973

- Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain: a journal of neurology*, *106* (Pt 3), 623–642.
- Maasen, S., Prinz, W., & Roth, G. (2003). *Voluntary action: brains, minds, and sociality*. New York, NY US: Oxford University Press.
- Malle, B. F. (1999). How people explain behavior: A new theoretical framework. *Personality and Social Psychology Review*, *3*, 23–48.
- Malle, B. F. (2006). Of windmills and strawmen: Folk assumptions of mind and action. In S. Pockett, W. P. Banks, & S. Gallagher (Eds.), *Does consciousness cause behavior? An investigation of the nature of volition* (pp. 207–231). Cambridge, MA: MIT Press.
- Malle, B. F. (2008). The fundamental tools, and possibly universals, of social cognition. In R. M. Sorrentino & S. Yamaguchi (Eds.), *Handbook of motivation and cognition across cultures* (pp. 267–296). New York, NY: Elsevier/Academic Press.
- Malle, B. F., Guglielmo, S., & Monroe, A. E. (2012). Moral, cognitive, and social: The nature of blame. In J. P. Forgas, K. Fiedler, & C. Sedikides (Eds.), *Social thinking and interpersonal behavior* (pp. 313–331). Philadelphia, PA: Psychology Press.
- Malle, B. F., & Holbrook, J. (2012). Is there a hierarchy of social inferences? The likelihood and speed of inferring intentionality, mind, and personality. *Journal of Personality and Social Psychology*, *102*, 661–684. doi:10.1037/a0026790
- Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, *33*, 101–121. doi:10.1006/jesp.1996.1314
- Malle, B. F., Moses, L. J., & Baldwin, D. A. (2001). *Intentions and intentionality: Foundations of social cognition*. Cambridge, MA: MIT Press.

- Mele, A. R. (2001). Acting intentionally: Probing folk notions. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition*. (pp. 27–43). Cambridge, MA: The MIT Press.
- Miller, J. S., & Feltz, A. (2011). Frankfurt and the folk: an experimental investigation of Frankfurt-style cases. *Consciousness and cognition*, *20*, 401–414.
doi:10.1016/j.concog.2010.10.015
- Monroe, A. E. (2013). *Blame as a function of choice capacity*. Unpublished raw data.
- 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*, 211–224. doi:10.1007/s13164-009-0010-7
- Montague, R. P. (2008). Free will. *Current Biology*, *18*, R584–R585.
- Nahmias, E. (2006). Folk Fears about Freedom and Responsibility: Determinism vs. Reductionism. *Journal of Cognition and Culture*, *6*, 215-237. doi:
10.1163/156853706776931295
- 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.
doi:10.1111/j.1475-4975.2007.00158.x
- Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2005). Surveying freedom: Folk intuitions about free will and moral responsibility. *Philosophical Psychology*, *18*, 561–584. doi:10.1080/09515080500264180
- Nichols, S. (2011). Experimental philosophy and the problem of free will. *Science*, *331*, 1401–1403. doi:10.1126/science.1192931

- Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous*, *41*, 663–685.
- Ohtsubo, Y. (2007). Perceived intentionality intensifies blameworthiness of negative behaviors: Blame-praise asymmetry in intensification effect. *Japanese Psychological Research*, *49*, 100–110. doi:10.1111/j.1468-5884.2007.00337.x
- Reeder, G. D. (2009). Mindreading: Judgments about intentionality and motives in dispositional inference. *Psychological Inquiry*, *20*, 1–18. doi:10.1080/10478400802615744
- Rigoni, D., Kühn, S., Gaudino, G., Sartori, G., & Brass, M. (2012). Reducing self-control by weakening belief in free will. *Consciousness and Cognition*, *21*, 1482–1490. doi:10.1016/j.concog.2012.04.004
- Roskies, A. L., & Malle, B. F. (2013). A Strawsonian look at desert. *Philosophical Explorations*, *online*, 1–20. doi:10.1080/13869795.2013.787439
- Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*, 299–309. doi:10.1016/S1364-6613(00)01506-0
- Schooler, J. W., Vohs, K. D., Nahmias, E., & Nadelhoffer, T. (2013, January). *Why does discrediting free will affect behavior?* Presented at the Society for Personality and Social Psychology, New Orleans, LA.
- Schurger, A., Sitt, J. D., & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences*, *109*, E2904–E2913. doi:10.1073/pnas.1210467109
- Stillman, T. F., & Baumeister, R. F. (2010). Guilty, free, and wise: Determinism and psychopathy diminish learning from negative emotions. *Journal of Experimental Social Psychology*, *46*, 951–960. doi:10.1016/j.jesp.2010.05.012

- 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*, 381–394. doi:10.1080/09515089.2011.556607
- Trevena, J., & Miller, J. (2010). Brain preparation before a voluntary action: Evidence against unconscious movement initiation. *Consciousness and Cognition, 19*, 447–456. doi:10.1016/j.concog.2009.08.006
- Vohs, K. D., & Schooler, J. W. (2008). The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychological Science, 19*, 49–54. doi:10.1111/j.1467-9280.2008.02045.x
- Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.
- Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1–34. doi:10.1016/S0010-0277(98)00058-4
- Young, L., & Saxe, R. (2009). Innocent intentions: A correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia, 47*, 2065–2072. doi:10.1016/j.neuropsychologia.2009.03.020

Acknowledgements

This chapter was made possible through the support of a grant from the Big Questions in Free Will project, funded by the John Templeton Foundation. The opinions expressed here are our own and do not necessarily reflect the views of the John Templeton Foundation.

Endnotes

¹ If some scholars then want to add statements about what free will “really” is, then they should no longer make claims about ordinary people but need to establish independent criteria for the truth of those statements.

² We report here the results for intentional behaviors only because there is obviously no free will, intentionality, and the like present for unintentional behaviors. These behaviors were included primarily to make it impossible to assume by default that every presented behavior is intentional, free, etc.

³ Choice and intentionality are not sufficient because an act chosen under severe constraints is not seen as free (Monroe & Malle, 2010).

⁴ Though each person received only one blameworthy behavior, we varied the blameworthy behaviors between subjects so that, between participants we collected data on six different blameworthy behaviors. Pretesting established that two behaviors were weakly negative (starting a fight with a coworker; sunbathing nude), two were moderately negative (stealing expensive clothes; tricking a cashier into giving too much change back), and two were strongly negative (cheating on a spouse; selling drugs to children). Following from Nichols and Knobe (2007), we hypothesized that the negativity of the behaviors might moderate the effect of threatening free will beliefs—stronger effects of the free will manipulation on weakly negative behaviors compared to strongly negative behaviors. However, we found no significant interaction between condition and behavior negativity ($p > .5$).