Bringing free will down to Earth: People's psychological concept of free will and its role in moral judgment

Andrew E. Monroe [a,*], Kyle D. Dillon [b], Bertram F. Malle [c]

[a] Florida State University, USA
[b] Harvard University, USA
[c] Brown University, USA

A B S T R A C T

Belief in free will is widespread, and this belief is supposed to undergird moral and legal judgment. Despite the importance of the free will concept, however, there remains widespread confusion regarding its definition and its connection to blame. We address this confusion by testing two prominent models of the folk concept of free will—a metaphysical model, in which free will involves a soul as an uncaused "first mover," and a psychological model, in which free will involves choice, alignment with desires, and lack of constraints. We test the predictions of these two models by creating agents that vary in their capacity for choice and the presence of a soul. In two studies, people's judgments of free will and blame for these agents show little to no basis in ascriptions of a soul but are powerfully predicted by ascriptions of choice capacity. These results support a psychological model of the folk concept of free will.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Ask people on the street if they have free will, and most will affirm they do. Beliefs about free will have a profound impact on social life, influencing moral judgment (Nichols, 2011), legal responsibility (Greene & Cohen, 2004; Krueger, Hoffman, Walter, & Grafman, 2013), job performance (Stillman et al., 2010), cheating behavior (Vohs & Schooler, 2008), aggression and helping (Baumeister, Masicampo, & DeWall, 2009), and even premotor neural activation (Rigoni, Kühn, Gaudino, Sartori, & Brass, 2012).

The study of free will—specifically, the study of people's belief in free will—has garnered widespread attention over the past decade. Yet for such an important concept, there remains widespread disagreement over the definition of free will (e.g., Nichols, 2011; Wegner, 2002) and over its relationship with such important facets of social life as morality (e.g., Nahmias, Morris, Nadelhoffer, & Turner, 2005; Nichols & Knobe, 2007). We therefore set out to untangle the definitional confusion over the folk concept of free will and assess the exact link between free will and moral blame. In particular, we compare two possible models of people's concept of free will and its relation to morality and provide two experiments that test the validity of these two models.

---

* Corresponding author. Address: Department of Psychology, Florida State University, 1107 West Call Street, Tallahassee, FL 32036, USA.
   E-mail address: Monroe@psy.fsu.edu (A.E. Monroe).

1.1. Two views of people's concept of free will

One prominent model of how ordinary people conceptualize free will argues that people are committed to substance dualism (Bloom, 2007), which entails (among other things) the concept of a soul as a "first mover" or "uncaused cause" for free will. People's concept of free will "is the idea that we make choices and have thoughts independent of anything remotely resembling a physical process" (Montague, 2008, p. R584), "laden with the concept of a soul, a non-physical, unfettered, internal source of choice-making" (Bargh & Earp, 2009, p. 13). On this view, people conceptualize human free will and agency as distinct from other types of physical systems.

For example, one study by Nichols (2004) asked groups of children and adults to reason about whether two types of events "had to happen." One was a physical event: a pot of water heating over a flame. The other was a choice event: a woman choosing to have vanilla ice cream. For both the physical and the choice events, Nichols asked his participants: "If everything in the world was the same right up until she chose vanilla [the water boiled], did Joan have to choose vanilla [did the water have to boil]?" (2004, p. 487). Participants were much more likely to agree that the water had to boil than to agree that the agent had to choose the vanilla ice cream.

Nichols (2004) took these data as evidence that people are committed to an indeterministic conception of human free will; however, others have gone farther to suggest people's folk concept of free will is inherently dualistic. "Common sense tells us that we exist outside of the material world—we are connected to our bodies and our brains, but we are not ourselves material beings, and so we can act in ways that are exempt from physical law" (Bloom, 2012, p. 1). This view that people's concept of free will is inextricably bound to concerns about indeterminism on the one side and requiring a soul as an "uncaused causer" on the other side is further buttressed by research showing that metaphysical considerations, such as highlighting the presence of a deterministic universe, which some take to be at odds with the folk view, can alter people's willingness to attribute free will and blame to agents (See Knobe, 2014; Nichols & Knobe, 2007; Roskies & Nichols, 2008).

While the characterization of people's folk concept of free will as metaphysical dualism has largely held sway in scholarly and popular circles (e.g., Bargh, 2008; Bargh & Earp, 2009; Bayer, Ferguson, & Gollwitzer, 2003; Bloom, 2007, 2012; Cashmore, 2010; Chisholm, 1966; Montague, 2008), some writers have recently noted that ordinary people's concept may actually resemble compatibilist theories of free will (Kane, 2011; Mele, 2014; Nahmias, Morris, Nadelhoffer, & Turner, 2006; Nahmias et al., 2005; Woolfolk, Doris, & Darley, 2006). So what *is* the folk concept of free will?

Monroe and Malle (2010) recently offered an empirically based model of the folk concept of free will. According to their model, people have a psychologically tractable concept of free will that is defined in terms of choice, alignment with one's desires, and freedom from constraints. Further, in contrast to the metaphysical characterization of people's concept of free will, concerns about souls, uncaused causes, and determinism are inconsequential for this psychological concept of free will.

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" (p. 214). This approach mirrored Malle and Knobe's (1997) research, which successfully demonstrated that the criteria for a folk concept such as intentionality could be elicited by directly asking people about the concept (e.g., "When you say that somebody performed an action intentionally, what does this mean?" p. 106). Supporting the psychological model of free will, the convergent meaning of free will was the ability to make a *choice* in line with one's *desires*, and *free of constraints*. Notably, not a single person mentioned the involvement of a soul or indeterminism.

In a congruous analysis of free will, Stillman, Baumeister, and Mele (2011) asked participants to produce an autobiographical account of actions they felt were either "of their own free will" or "not the result of free will." Participants in the "free will" condition reported behaviors associated with pursuing desired goals, making choices, and acting against external forces (e.g., temptation or pressure from others), whereas participants in the "no free will" condition wrote about behaviors in the presence of powerful authority figures. Together these studies present an emerging picture of a lay view of free will that is psychological—defined by choice, alignment with desires, and freedom from constraints.

1.2. Free will and morality

For many scholars, questions concerning free will carry great import because free will 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. Darwin put the sentiment succinctly, arguing that without a belief in free will "one deserves no credit for anything... nor ought one to blame others" (Darwin, 1840, p. 27). Neither of the two major models of people's folk concept of free will disputes the importance of free will for morality. However, the metaphysical and psychological models of people's concept of free will differ sharply regarding the criteria that are necessary for agents to have free will, and therefore to be morally responsible for their actions. The metaphysical model argues that for an agent to have free will and to be morally blameworthy, (1) the agent must have a soul; and (2) in virtue of having a soul the agent is able to uniquely intervene on the world, free of deterministic causes (i.e., the soul represents a "first mover" or "uncaused cause"). The psychological model, by contrast, proposes that free will and moral blame require that (1) an agent has the capacity for choice and intentional action; and (2) the agent is (reasonably) free from constraints.

1.3. Disentangling souls from agents

In ordinary human agents, ascriptions of souls, indeterminism, choice, and intentional agency are typically confounded. To disentangle these properties, and to test the relationship between judgments of free will and morality with souls and indeterministic causation on the one hand and with ascriptions of choice and intentionality on the other hand, we conducted two studies. In the first study we constructed five different agents (e.g., human, robot, cyborg) whose descriptions varied in a number of features, most notably in either having or lacking a human brain, 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, indeterministic causal powers, choice, and free will. In Study 2 we sharpened our test of the metaphysical and psychological models of free will by constructing agents that varied specifically in their capacity for choice on the one hand and having a soul on the other hand. We then examined whether choice capacity or having a soul better predicted free will and blame judgments.

Thus, the two studies test the key predictions of both the metaphysical and psychological models of free will by tracking the relationships of multiple judgments across a variety of agents. The wide-spread claim that people are committed to the presence of an uncaused soul in their free will concept would be supported if, in people's patterns of judgments, having a soul predicts having free will and if having free will (along with a soul) predicts moral judgments. A psychological model such as Monroe and Malle's (2010) would be supported if having a capacity for choice predicts free will as well as moral judgments, while the soul lacks predictive power.

2. Study 1

Study 1 introduced five agents who varied in the presence of a human brain, physiology, emotions, and—most important—choice capacity. Participants ascribed various capacities to the agent, including having a soul and having free will, and made moral judgments about the agent's norm-violating behaviors.

2.1. Methods

2.1.1. Participants

A total of 221 participants from Amazon Mechanical Turk (AMT) began the online study. Prior to data analysis we set three criteria for excluding participants from analyses. Participants were excluded if they: (1) failed to complete the entire study; (2) failed *both* catch questions; or (3) spent an implausibly low amount of time reading the agent description or blame questions. In total, 24 participants met at least one of these criteria and were therefore omitted from analyses (final $N = 197$).

The remaining participants were 60% female and predominantly White (76%), with an average age of 35.5 years ($SD = 13.6$) and a broad range of education: 26% reported finishing only high school; 61% attained a 2- or 4-year college degree; and 12% attained a graduate degree. Political attitudes were slightly liberal, $M = 5.9$, $SD = 1.82$, on a 1 (extremely conservative) to 9 (extremely liberal) scale. Religiousness[1] averaged $M = 1.6$ ($SD = 1.40$) on a 0–4 scale (0 = not at all religious; 4 = very religious).

2.1.2. Procedure and materials

Participants were randomly assigned to read a description (see Supplementary material) of one of the following agents:

(1) a normal human ($N = 33$);
(2) an "akratic" human ($N = 37$), who had an "inability to use his thoughts to control his actions";
(3) a cyborg—a human brain in a robot body ($N = 43$);
(4) an AI in a human body ($N = 43$);
(5) an advanced robot ($N = 41$).

After reading the agent description, participants judged how much blame the agent deserved for seven norm-violating actions (see Supplementary materials) on a 0 (no blame at all) to 6 (the most blame you would ever give) scale. The seven items showed substantial internal consistency ($\alpha = .95$) and were averaged into a single blame rating.

Following the blame judgments, participants responded to six questions about the particular agent's capacities, using a −3 (definitely no) to +3 (definitely yes) scale with a midpoint of 0 (not sure): (1) Can the agent intervene on the normal causal flow of the universe? (2) Was the agent the sole cause of his action? (3) Does the agent have the capacity for choice? (4) Does the agent have free will? (5) Can the agent act intentionally? (6) Does the agent have a soul?

---

[1] Participants' religiousness did not significantly affect free will or blame judgments, nor did it interact with ascriptions of choice or souls ($ps > .1$, see Supplementary materials).
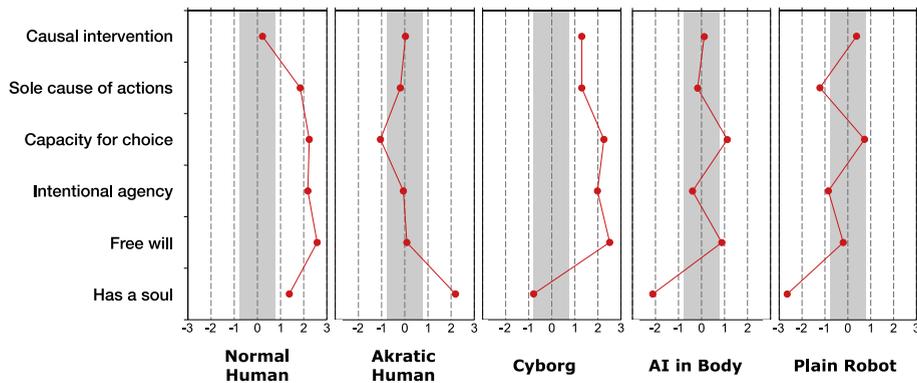
Fig. 1. Profiles of ascribed capacities for each of five agents. All ascriptions were made on scales ranging from "Definitely no" (−3) through "Not sure" (0) to "Definitely yes" (+3). For each agent, points outside the gray area differ significantly from the 0 point ($p < .05$, two-tailed).

2.2. Results

2.2.1. Agents' properties

Fig. 1 displays the results of capacity ascriptions across the five agents. Two findings stand out. First, the akratic human was denied the capacity for choice and had doubtful intentional agency and free will, even though he was clearly granted a soul. Second, people granted the cyborg capacities for choice, intentional agency, and free will, while simultaneously denying him a soul. Thus, once people have an opportunity to independently ascribe souls and free will to agents, the two capacities break apart.

2.2.2. What constitutes free will?

We conducted multiple regressions of free will on capacities for each agent type separately and for the entire sample. We report both unique and shared predictive variance components (semi-partial r^2) because, due to considerable correlations among predictors, unique components accounted for only a limited portion of the predictive variance.[2] Moreover, shared predictive variance reveals which capacities hang together in constituting free will ascriptions.

For the normal human, free will ascriptions were predicted well by capacities (overall $R^2 = .87$). The only unique predictive variance components were choice ($r^2 = .02$, $p = .04$) and intentionality ($r^2 = .04$, $p = .01$). The dominant shared-variance contributions came from choice with intentionality ($r^2 = .36$, $p < .001$) and from these two with sole-cause ($r^2 = .38$, $p < .001$), as well as smaller shared variance among all the aforementioned capacities and soul ($r^2 = .07$), $F(1, 25) = 14.1$, <.001, but no unique contribution from soul itself. When applying backward elimination to simplify the model (F-to-remove criterion: $p > .10$), only choice and intentionality survived, with an R^2 of .87.

For the akratic human, overall R^2 was .52. The only significant contributions came from the unique components of choice ($r^2 = .07$) and intentionality ($r^2 = .06$) and their shared variance ($r^2 = .26$), adding up to an r^2 of .38 ($p < .001$). All other predictors (unique or shared) were small and nonsignificant ($p > .13$). Under backward elimination, only choice and intentionality survived, with an R^2 of .44.

For the cyborg, overall R^2 was .63. Choice and intentionality again dominated the prediction, with their unique and shared contributions adding up to $r^2 = .18$ ($p < .001$) and additional predictive variance between them and sole cause ($r^2 = .22$, $p < .001$), whereas sole cause alone made no contribution ($r^2 < .01$). The causal flow variable shared a small contribution with all three of these predictors ($r^2 = .05$, $p = .03$) but also made no unique contribution ($r^2 < .01$). Under backward elimination, choice and intentionality survived, and soul now made a small unique contribution ($r^2 = .04$) to an R^2 of .62.

The AI in a human body was less predictable (overall $R^2 = .32$), with free will ratings lying mostly in the "not sure" range. Still, choice and intentionality again contributed through unique and shared components ($r^2 = .15$, $p = .04$). No other components were significant; nonsignificant contributions came from shared variance between intentionality and sole cause ($r^2 = .05$, $p = .14$) and between intentionality, choice, and causal flow ($r^2 = .05$, $p = .11$). Because intentionality was involved in all these components, it was the only one that survived backward elimination, with an R^2 of .26.

For the advanced robot, overall R^2 was .57, predicted from a combination of unique and shared components of choice, intentionality, and sole cause, adding up to an r^2 of .42 ($p < .001$). Smaller contributions came from having a soul ($r^2 = .06$, $p = .05$) and shared variance between intentionality and having a soul ($r^2 = .06$, $p = .04$). However, the involvement of soul was entirely due to a single participant who gave the advanced robot a score of +2 on the soul variable whereas everybody else gave the advanced robot scores of −1 to −3. After removing this participant, the overall R^2 was still .53, mainly comprising unique and shared components of choice, intentionality, and sole cause (r^2 of .49, $p < .001$). The contributions of soul

---

[2] We report β values, which indicate unique contributions, in the Supplementary material.
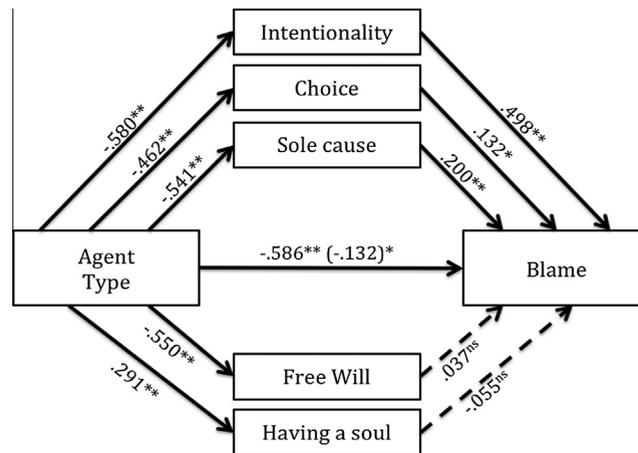
Fig. 2. Ascriptions of intentionality, choice, and sole cause largely mediate the impact of agent type on blame ($\beta = -.586$ reduced to $-.132$). $^*p < .05$, $^{**}p < .001$.

(shared or unique) dropped below 1%. Under backward elimination, intentionality and sole cause survived, with an overall R^2 of .50.

In the entire sample, similar results obtained, but because of the large sample size ($N = 184$), many small variance components (down to 1%) were significant at $p < .05$. Predictively dominant were the unique and shared components of choice and intentionality ($r^2 = .25$), along with their shared variance with sole cause ($r^2 = .20$). Small contributions came from the unique component of soul ($r^2 = .04$) and its shared components with intentionality and sole cause ($r^2 = .05$). Under backward elimination, all but causal intervention survived, but intentionality, choice, and sole cause made up $r^2 = .53$ out of an overall R^2 of .66.

Thus, a cluster of capacities constitutes free will: choice, intentionality, and (to a lesser extent) sole cause; having a soul plays virtually no role. Because the meaning of "sole cause of one's action" is somewhat ambiguous, we report shortly a follow-up study to clarify its meaning. First, however, we report on the relationship between blame and free will.

2.2.3. Does blame rely on free will?

The agent manipulation had a substantial impact on people's blame judgments, $F(4, 192) = 27.42$, $p < .001$, $\eta^2 = .36$. Tukey post hoc tests revealed that the normal human ($M = 5.4$) and the cyborg ($M = 5.1$) received the highest amounts of blame; the AI/human body hybrid received less blame ($M = 3.9$) than each of them, $ps < .001$, $ds = 1.27$ and 0.99, respectively; and the akratic human ($M = 3.0$) and advanced AI robot ($M = 2.8$) received yet lower amounts of blame, $ps = .039$ and .004, $ds = 0.57$ and 0.63, respectively.

More important, variability in blame was highly predictable from specific agent capacities, $R^2 = 0.66$, $F(5, 181) = 74.0$, $p < .001$. Because of high correlations among the predictors we aggregated choice and intentionality into an "agency" variable. In the full sample, the dominant predictors were the unique component of this agency variable ($r^2 = .14$), its shared component with free will ($r^2 = .14$), and their jointly shared variance with sole cause ($r^2 = .24$). All other contributions were $r^2 < .03$. Under backward elimination, only agency and sole cause survived (free will made no unique contributions), with the model's overall R^2 remaining at 0.66.

We also performed a multiple mediation analysis of the impact of agent type on blame (initial $\beta = -.59$). This impact dropped to $\beta = -.13$ when taking into account the predictive power of ascribing intentionality, choice, and sole cause, $z = -8.80$, $p < .001$ (Preacher & Hayes, 2008). In contrast to the substantial mediation of these three properties, having free will[3] or having a soul made no unique contribution to the prediction of blame (see Fig. 2).

3. Follow-up study

These results offer initial support for a psychological folk theory of free will. However, the "sole cause of action" capacity is ambiguous. It may be interpreted as "free from outside causal influence, free from constraints" or as a metaphysical assumption about an agent being an "uncaused cause." To resolve this ambiguity we performed a follow-up study replicating Study 1's "Normal human" condition, with one change: After answering the sole cause question, participants were asked to explain in a few sentences how they understood the question.

---

[3] However, when choice, intentionality and sole cause were omitted from the model, free will significantly mediated judgments of blame, $z = -5.93$, $p < .001$.

3.1. Methods

3.1.1. Participants and procedure

The sample ($N = 46$) was comparable to that of Study 1. The sample was similar in age ($M = 35.6$ years, $SD = 13.4$), gender distribution (43% female), ethnic composition (87% white), and again represented a diverse range of education: 26.1% ($n = 12$) reported finishing high school as their highest level of education; 19.6% ($n = 9$) attained a 2-year degree; 45.7% ($n = 21$) attained a 4-year degree, and 6.5% ($n = 3$) attained a Master's degree or higher. Participants in this study were slightly less liberal ($M = 5.76$, $SD = 1.72$) and also slightly less religious ($M = 1.10$, $SD = 1.21$) than in Study 1.

Participants were recruited from AMT and completed a replication of Study 1's "Normal human" condition, with one important change: After answering the sole cause question participants responded to an open-ended question: "How did you understand the question above ('Was Norm the sole cause of his action')? Rephrase the question in your own words."

3.2. Results

The 46 participants provided 67 codeable responses. These responses were content coded by the authors after reading through a portion of the responses and identifying seven major coding categories: (1) lack of (internal or external) constraints, (2) making a choice, (3) being responsible for his actions, (4) being the only reason for the outcome, (5) having control over his actions, (6) having free will, and (7) metaphysical explanations. 93% of responses could be classified into these categories. Each response was permitted to be coded into multiple categories, but only three were. Inter-rater reliability between the coders was high ($\kappa = .82$; 85% agreement).

This classification of responses showed that 97% of people interpreted the sole cause question as asking about psychological limitations and inputs to an agent's actions. The single most frequent interpretation (given by 45% of participants) was that the "sole cause" question asked about the presence or absence of strong constraints (e.g., "Did anybody else have influence over Norm's decisions?"). The remaining interpretations focused on the agent making his own choice (given by 28% of participants), being the one who is responsible for his actions (28%), being the only reason for the actions (15%), and being in control of his actions (13%).[4] Metaphysical interpretations of the question, by contrast, were rare—only three of the 46 participants interpreted the question in such terms (e.g., "Do we control our actions or does the universe?").

4. Study 2

The preceding results provide evidence against the metaphysical model of the folk concept of free will and favor the psychological model. However, a critique of Study 1 might be that correlating ascriptions of the agents' capacities (e.g., soul and choice) with concurrent judgments of free will and blame offers only indirect evidence about the validity of these two models. In Study 2 we orthogonally manipulated two critical factors—having a soul and the capacity for choice—that underpin the metaphysical and psychological models of people's folk concept of free will, respectively, and we let them compete against each other in predicting ascriptions of free will and judgments of blame. In addition, we measured people's belief in the existence of a soul and their religiousness as possible predictors of those judgments.

4.1. Method

4.1.1. Participants

A total of 182 participants from AMT began the online study. Participants were excluded from the final data set using the same exclusion criteria as Study 1 with one addition: Participants were excluded from the data set if they had previously participated in Study 1. Thirty participants met this criterion and were omitted. Of the 152 participants who had not participated in Study 1, 16 participants failed *both* of the catch questions, and 12 participants had implausibly low reading times for the agent description ($n = 9$) or blame questions ($n = 3$). In total, 28 participants met at least one of the Study 1 exclusion criteria and were therefore omitted from analyses (final $N = 124$).

Participants were predominantly female (65%) and White (82%) with an average age of 28.2 years ($SD = 9.45$) and a broad range of education as in Study 1. Participants were slightly more conservative than in Study 1 ($M = 3.1$, $SD = 1.54$) and again moderately religious ($M = 2.10$, $SD = 1.27$).

4.1.2. Procedure and materials

Participants read one of four agent descriptions that pretesting confirmed to represent the cells of a 2 (choice capacity present/absent) × 2 (soul present/absent) between-subjects design: Normal human (choice present/soul present), Cyborg (choice present/soul absent), Akratic Human (choice absent/soul present), and Robot (choice absent/soul absent). As in Study 1, participants made blame judgments (for five norm-violating actions), followed by ratings of the agent's capacities, which were identical to those used in Study 1 and were presented in a randomized order.

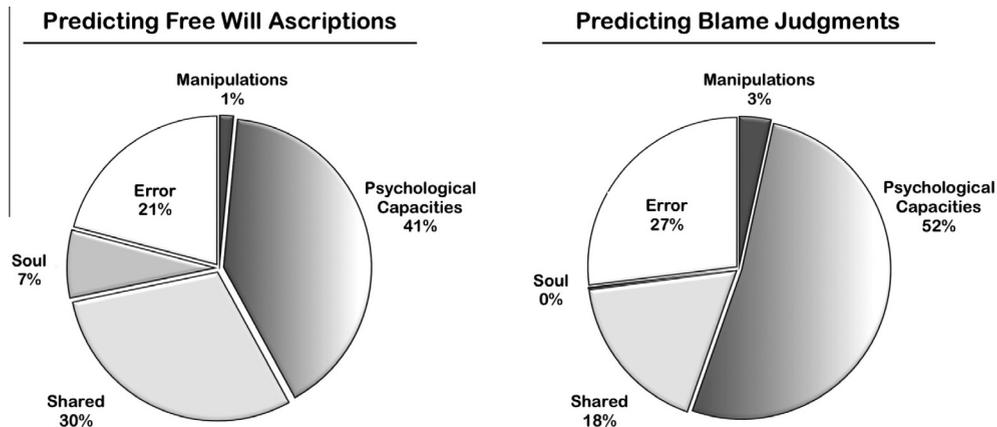[4] Percentages do not sum to 100 because some people provided response in more than one category.

Fig. 3. Proportion of variance in free will and blame judgments explained by each predictor. Predictive variance labeled "Shared" is shared between the agent manipulations, the psychological capacities, and soul ascriptions.

In addition, we assessed participants' personal belief in the existence of a soul with a yes–no question: "Do you believe in the existence of a soul?" The majority of participants (68%) reported believing in a soul. However, participants' reported belief in a soul did not significantly affect free will judgments or blame judgments, nor did it interact with the choice and soul manipulations ($ps > .1$).

4.2. Results

We performed two hierarchical multiple regressions, one for free will ascriptions and one for blame judgments. The two manipulated factors (soul present/absent, choice present/absent) were always entered first, followed by the "psychological" variables (choice, intentionality, sole cause), and finally the "metaphysical" variables (soul, causal intervention). Because of correlations of .58 to .83 among the three psychological predictors we aggregated them into a single "psychological capacities" variable.[5] Causal intervention never had any predictive power, so we eliminated it. We also tested religiousness and belief in souls, but neither made any contributions ($r^2 < .01$, $p > .20$). We again report unique and shared predictive variance components.

4.2.1. Free will ascriptions

In Step 1, the manipulated factors explained 60% of variability in free will ascriptions, $F(2,120) = 191.8$, $p < .001$. In Step 2, their direct predictive power was greatly reduced ($r^2 = .05$, $p < .001$) once psychological capacities were included in the model. These capacities explained 70% of the free will variance, $F(3,119) = 113.8$, $p < .001$, largely generated by the manipulations (shared $r^2 = .54$). In Step 3, soul ascriptions were entered, which predicted 7% of free will variability ($p < .001$), of which $r^2 = .05$ was generated by the soul manipulation. As a result, unique effects of manipulations were further reduced ($r^2 = .015$, $p = .02$). After Step 3, the psychological capacities were still strongly involved in predicting free will variance overall ($r^2 = .70$, $p < .001$), but 30% explained variance was now shared—a small amount with soul ascriptions ($r^2 = .04$, $p < .001$) and a considerable amount with both soul ascriptions and the two manipulations ($r^2 = .26$, $p < .001$). Thus, the two agent manipulations created a certain amount of convergence among ascriptions of psychological capacities and ascriptions of having a soul, and these converging judgments partially predicted free will. The final model, displayed in Fig. 3 (left panel), amounted to $R^2 = .79$, $F(4,118) = 126.7$, $p < .001$.

4.2.2. Blame judgments

In Step 1, the manipulated factors explained 58% of variability in blame judgments, $F(2,121) = 82.4$, $p < .001$. In Step 2, with psychological capacities and free will ascriptions entered, the manipulations' direct predictive power was reduced to $r^2 = .03$ ($p = .002$). However, the overall model improved to $R^2 = .73$, $F(4,118) = 78.9$, $p < .001$, with substantial prediction involving psychological capacities ($r^2 = .67$, $p < .001$). This prediction comprises unique variance (6%), some variance generated by the manipulations (8%), but also predictive variance shared with free will ascriptions (7%) and especially such shared variance generated by the manipulations (48%). Importantly, however, free will ascriptions on their own did not predict blame ($r^2 < .01$, $p > .5$) and therefore could be left out of the model without losing predictive power. Apparently, the two agent manipulations had a strong joint influence on perceived psychological capacities and ascribed free will, which is understandable given that these capacities successfully predicted free will ascriptions, as shown in the previous section.

---

[5] Models that include only one of the three predictors are almost as strong as the model reported here.

In Step 3, soul ascriptions were entered, but the model did not improve. Fig. 3 (right panel) shows that the manipulations' direct predictive power remained at 3% and psychological capacities continued to be involved in virtually the entire prediction ($r^2 = .69$, $p < .001$). Although 38% of this predictive variance was shared with free will and 18% with soul ascriptions, neither soul ascriptions nor, as before, free will ascriptions made any contribution independent of psychological capacities ($r^2 < .01$, $ps > .5$). As a result, a model with just the manipulated factors and the aggregate psychological capacity variable as predictors was strong ($R^2 = .72$) and statistically indistinguishable ($p > .20$) from the best model including other predictors ($R^2 = .73$).

4.3. Discussion

The manipulations of perceived choice capacity and having a soul had substantial effects on corresponding capacity judgments, and these judgments directly predicted people's ascriptions of free will and blame. In particular, a bundle of psychological capacities that includes intentionality, choice, and being the sole cause of one's action (earlier shown to be understood as absence of constraints) strongly predicted free will and blame judgments. Free will ascriptions were predominantly explained by these psychological capacities, with a small unique influence of soul ascriptions, and variance shared between the two. Blame judgments were also primarily a function of perceived psychological capacities, with no unique contribution from soul ascriptions but some variance shared between the two.

What can we make of this shared variance between psychological capacities and soul ascriptions (generated primarily by the agent manipulations)? One possibility is that people determine an agent's capacities such as intentionality in part by asking themselves whether the agent has a soul. This seems rather unlikely, however, given the vague meaning of "soul" on the one hand and the clear folk meaning of intentionality and choice on the other hand (Malle & Knobe, 1997), which is also reflected in early conceptual development (Malle, Moses, & Baldwin, 2001) and speedy judgments of intentionality (Malle & Holbrook, 2012). The opposite direction—that at least some people would judge whether an agent has a soul by asking themselves whether the agent has certain cognitive capacities—seems more likely. We have tentative evidence for such a relationship from interviews with community members, who were asked to define what "having a soul" means to them (Monroe & Malle, in press). About a third of participants likened the soul to the self or one's identity; another third described it as some sort of spirit or life force; and one third defined it as a cognizer or decision maker. For example, one person said: "it's a mental concept, but it's also sort of the innate part of a person that… makes decisions that direct… their actions."

Finally, a more statistical explanation for the relationship between psychological capacity ascriptions and soul ascriptions (and their joint relationship with free will or blame) is that in the case of the plain robot, people decisively denied the agent a soul, intentionality, free will, and eligibility for blame, which created a strong point swarm of low-low scores in whole-sample correlations among these variables. Indeed, when we reran the multiple regression of free will ascriptions on all candidate predictors but omitting the data from the plain robot (leaving the normal human, akratic human, and cyborg), psychological capacities explained 50% of free will variability, soul ascriptions explained 7%, and no shared variance between them was explanatory. This suggests that soul and psychological capacities converge only when they are all seen as absent, as in a robot.

5. Conclusions

We tested two models of the folk concept of free will and assessed connections between free will and blame. We offer two main conclusions. First, the present studies contradict a metaphysical model of the folk concept of free will (e.g., Bargh & Earp, 2009; Cashmore, 2010; Montague, 2008). Even though people ascribed souls to some agents and not others, having a soul was only weakly related to having free will and not at all to blame, and this was true regardless of whether people expressed believing in souls or not. Additionally, while soul ascriptions explained a statistically significant portion of free will variance, this effect was small (4% in Study 1, 7% in Study 2). By contrast, psychological capacities (primarily choice and intentionality) explained 53% and 41% of the variance in free will judgments. These findings, coupled with the absence of predictive power of variables assessing nondeterminism ("causal intervention in the universe") and religiosity, cast serious doubt on the metaphysical model of people's concept of free will and instead support a psychological model of free will (Monroe & Malle, 2010). However, we have not yet comprehensively tested potential beliefs in nondeterministic causation (e.g., Ebert & Wegner, 2011). We expect challenges of measurement here, because it is not entirely clear how such causation is expressed in everyday language. "Breaking the causal flow of the universe" was our initial attempt, but people may not have fully understood this phrase, and we intend to experiment with other formulations in the future.

The second conclusion from these studies is that judgments of blame did not at all rely on the notion of a soul and, surprisingly, also not on ascriptions of free will. Instead, blame judgments were grounded in the psychological capacities for choice, intentional action and, to a smaller extent, in an agent's unconstrained generation of action.

This second conclusion may be controversial. How can the assumption of free will be irrelevant for moral judgments? Previous research found that the folk concept of "free will" consists of three components: choice, alignment with one's desires, and absence of strong constraints (Monroe & Malle, 2010). The first two components are arguably captured in the present studies by the capacities of choice and acting intentionally (i.e., in line with one's desires), and the third component,

absence of constraints, is the dominant interpretation of the "sole cause" capacity (see follow-up study to Study 1). When these three components were entered as predictors of blame judgments in our studies, there was nothing left for an abstract free will capacity to predict. Though ascriptions of these psychological capacities and ascriptions of free will shared substantial variance, only the psychological capacities had unique predictive power for judgments of blame. That, we argue, is because people understand "free will" as a chosen, desired intentional action absent constraints (Monroe & Malle, 2010). There is no residual meaning of free will—metaphysical or otherwise—that has relevance for blame judgments.

Acknowledgments

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.concog.2014.04.011.

References

Bargh, J. A. (2008). Free will is unnatural. In J. Baer, J. C. Kaufman, & R. F. Baumeister (Eds.), *Are we free? Psychology and free will* (pp. 128–154). New York, NY: Oxford University Press.

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.

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. http://dx.doi.org/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.

Bloom, P. (2007). Religion is natural. *Developmental Science, 10*, 147–151. http://dx.doi.org/10.1111/j.1467-7687.2007.00577.x.

Bloom, P. (2012, March 18). Free will does not exist. So what? *The chronicle of higher education.*

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. http://dx.doi.org/10.1073/pnas.0915161107.

Chisholm, R. M. (1966). Freedom and action. In K. Lehrer (Ed.), *Freedom and determinism* (pp. 11–44). New York, NY: Random House.

Darwin, C. R. (1840). Old and useless notes about the moral sense and some metaphysical points. (P. H. Barrett, Trans.).

Ebert, J. P., & Wegner, D. M. (2011). Mistaking randomness for free will. *Consciousness and Cognition, 20*, 965–971. http://dx.doi.org/10.1016/j.concog.2010.12.012.

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. http://dx.doi.org/10.1098/rstb.2004.1546.

Kane, R. (2011). *The Oxford handbook of free will* (2nd ed.). New York, NY US: Oxford University Press.

Knobe, J. (2014). Free will and the scientific vision. In E. Machery & E. O'Neill (Eds.), *Current controversies in experimental philosophy*. Routledge.

Krueger, F., Hoffman, M., Walter, H., & Grafman, J. (2013). An fMRI investigation of the effects of belief in free will on third-party punishment. *Social Cognitive and Affective Neuroscience.* http://dx.doi.org/10.1093/scan/nst092.

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. http://dx.doi.org/10.1037/a0026790.

Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology, 33*, 101–121. dx.doi.org/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. (2014). Free will and substance dualism: The real scientific threat to free will. In W. Sinnott-Armstrong (Ed.). *Moral psychology, Free will and moral responsibility* (vol. 4). Cambridge, MA: MIT Press.

Monroe, A. E., & Malle, B. F. (in press). Free will without metaphysics. In A. R. Mele (Ed.), *Surrounding free will*. New York, NY: Oxford University Press.

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. http://dx.doi.org/10.1007/s13164-009-0010-7.

Montague, R. P. (2008). Free will. *Current Biology, 18*, R584–R585.

Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2005). Surveying freedom: Folk intuitions about free will and moral responsibility. *Philosophical Psychology, 18*, 561–584. http://dx.doi.org/10.1080/09515080500264180.

Nahmias, E., Morris, S. G., Nadelhoffer, T., & Turner, J. (2006). Is incompatibilism intuitive? *Philosophy and Phenomenological Research, 73*, 28–53. http://dx.doi.org/10.1111/j.1933-1592.2006.tb00603.x.

Nichols, S. (2004). The folk psychology of free will: Fits and starts. *Mind and Language, 19*, 473–502. http://dx.doi.org/10.1111/j.0268-1064.2004.00269.x.

Nichols, S. (2011). Experimental philosophy and the problem of free will. *Science, 331*, 1401–1403. http://dx.doi.org/10.1126/science.1192931.

Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous, 41*, 663–6859.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior research methods, 40*, 879–891. http://dx.doi.org/10.3758/BRM.40.3.879.

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. http://dx.doi.org/10.1016/j.concog.2012.04.004.

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

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. http://dx.doi.org/10.1080/09515089.2011.556607.

Stillman, T. F., Baumeister, R. F., Vohs, K. D., Lambert, N. M., Fincham, F. D., & Brewer, L. E. (2010). Personal philosophy and personnel achievement: Belief in free will predicts better job performance. *Social Psychological and Personality Science, 1*, 43–50. http://dx.doi.org/10.1177/1948550609351600.

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. http://dx.doi.org/10.1111/j.1467-9280.2008.02045.x.

Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.

Woolfolk, R. L., Doris, J. M., & Darley, J. M. (2006). Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition, 100*, 283–301. http://dx.doi.org/10.1016/j.cognition.2005.05.002.