Neuroscience and the soul: Competing explanations for the human experience

Jesse Lee Preston *, Ryan S. Ritter, Justin Hepler

Department of Psychology, University of Illinois at Urbana-Champaign, 603 E Daniel St., Champaign, IL, 61820, USA

Abstract

The development of fMRI techniques has generated a boom of neuroscience research across the psychological sciences, and revealed neural correlates for many psychological phenomena seen as central to the human experience (e.g., morality, agency). Meanwhile, the rise of neuroscience has reignited old debates over mind–body dualism and the soul. While some scientists use neuroscience to bolster a material account of consciousness, others point to unexplained neural phenomena to defend dualism and a spiritual perspective on the mind. In two experiments we examine how exposure to neuroscience research impacts belief in the soul. We find that belief in soul decreases when neuroscience provides strong mechanistic explanations for mind. But when explanatory gaps in neuroscience research are emphasized, belief in soul is enhanced, suggesting that physical and metaphysical explanations may be used reflexively as alternative theories for mind. Implications for the future of belief in soul and neuroscience research are discussed.

1. Introduction

Belief in the soul—a non-physical essence of a being—has been an important subject of philosophy and science for thousands of years (e.g., Plato, 2005; Descartes, 1641/2000). Some scholars have recently argued that belief in souls is culturally universal and hard-wired in cognitive processes used in agency detection (Bloom, 2004; Bering, 2006). But more than just a way to understand other minds, belief in the soul also helps people to explain the experience of their own mind. Whenever one thinks, feels emotion, or exercises free will, subjective experience seems to magically occur and is not obviously tied to any physical event (Wegner, 2003). The very act of introspection suggests a qualitative difference between the mental and the physical, and so it feels as though we are made of two parts: mind and body (Descartes, 1641; Ryle, 1949). Although the physical origin of the body is intuitively understood, the origin of the mind is less clear; indeed, the mind appears to arise from some extra-physical force, and the concept of the soul is commonly evoked as the source of this ineffable essence of self.

To the extent that belief in the soul is used as a metaphysical explanation for the mind, this belief may be threatened by physical explanations for the mind. The present research examines how belief in the soul is affected by neuroscience research that implies a physical origin of the mind. fMRI studies have uncovered neural correlates for many psychological phenomena seen as central to the human experience, including moral judgments (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001), emotion (LeDoux, 1996), and personal agency (Farrer & Frith, 2002). Accompanied by vivid images of the brain “lighting up” during mental activity, fMRI research appears to finally provide hard evidence that the mind is grounded in the physical. Moreover, the appeal of fMRI research extends beyond academia and has captured the attention of the general public. Laypersons express greater interest and belief in psychological research when it also contains neuroscience information (McCabe & Castel, 2008), even if that information does not provide additional support.
for the theory beyond what the behavioral data demonstrate (Weisberg, Keil, Goodstein, Rawson, & Gray, 2008). Widespread acceptance of fMRI as an explanatory tool may also have an important effect on general beliefs about the soul. Work on causal discounting demonstrates that alternate explanations for the same phenomenon can compete with each other on a cognitive level, such that increasing belief in one diminishes belief in the other (Morris & Larrick, 1995; Sloman, 1994). For example, reading scientific explanations for important phenomena (e.g., evolution) reduces belief in religious explanations (e.g., creationism) (Lawson & Weser, 1990; Shariff, Cohen, & Norenzayan, 2008), but when scientific explanations are framed as weak, they can actually bolster belief in supernatural explanations (Preston & Epley, 2009). We propose that a similar reflexive relationship may also occur for physical vs. metaphysical explanations of the mind, with implications for neuroscience research on the belief in the soul. As neuroscience continues to discover neural correlates of more and more essential psychological processes, the brain could edge out the soul as the prevailing explanation for the mind (Clark, 2010; Farah, 2005).

On the other hand, if neuroscience seems limited in its ability to explain psychological experiences, exposure to that research could enhance belief in the soul. Despite the many impressive breakthroughs of fMRI studies, there remains one epistemological issue of the mind that neuroscience may not be able to solve, dubbed the “hard problem of consciousness” (Chalmers, 1996) or the “explanatory gap” (Levine, 1983). In sum, although neuroscientists can identify neural correlates associated with mental processes, they are still unable to explain precisely how activity in the brain creates the experience of these mental phenomena. This issue can have some important implications for belief in the soul. If the neural activity captured by fMRI serves to demystify the mind, awareness of an explanatory gap may only re-mystify the mind. Indeed, while some scientists use neuroscience to bolster a material account of consciousness (e.g., Crick, 1994), others point to unexplained neural phenomena to defend dualism and a spiritual perspective on the mind (e.g. Schwartz, Stapp, & Beauregard, 2005). Whether a legitimate concern or not (for a discussion, see Dennett, 1991; Nagel, 1974), an apparent explanatory gap leaves some aspects of the mind unexplained and so re-opens the intuitive plausibility of metaphysical explanations.

1.1. Research overview

The present research examines how neuroscience explanations for psychological phenomena can impact lay belief in the soul, guided by two complementary hypotheses:

**H1.** Exposure to neuroscience research with strong mechanistic explanations for psychological experience will decrease belief in the soul as an alternative explanation for the mind.

**H2.** Exposure to neuroscience research with weak mechanistic explanations for psychological experience (i.e., research highlighting the explanatory gap) will increase belief in the soul as an alternative explanation for the mind.

### 2. Experiment 1

#### 2.1. Method

One hundred fifty-one undergraduates (95 women) volunteered to participate for partial course credit. Participants were randomly assigned to one of three conditions: psychology information, neuroscience information, or a control condition. All participants were told that they would complete three unrelated studies for credit. The first task was framed as a study to measure students’ judgments about college course descriptions. Subjects read brief course descriptions, selected the best title for the course between two options we provided, and rated their interest in the course. All participants first read a course description on introduction to geology, then a survey course on Shakespeare. In the psychology and neuroscience conditions, subjects then read two additional course descriptions: one on the study of love and one on morality. Both the Love and Morality course descriptions raised a number of questions about the psychological phenomenon (e.g., Why do people fall in love at first sight? Are some people inherently good or bad?). In the neuroscience condition, the descriptions also referred to recent evidence for the neural basis of love/morality (see Supplementary materials). The psychology description therefore raised the same questions regarding the phenomenology of the mind without providing any physical mechanism for the psychological experiences. As with the two previous course descriptions, subjects selected the best name for each course description between two options. For the Love course, two options were: “Mechanisms of Love” or “Mystery of Love”. For the Morality course, the two options were: “Moral Mechanisms” or “The Moral Compass”. These items therefore served as a manipulation check, as the first options suggested a mechanistic explanation for the phenomena, whereas the second options suggested the presence of some explanatory gap in the understanding of the phenomena.

After the course description task, participants completed some filler tasks (spatial reasoning items, and four 4 items taken from “Reading the mind in the eyes” test), which they were told would measure visual processing as the second part of the study.

#### 2.1.1. Body–soul dilemmas

In the third part of the study, participants played a game called “Staying Alive” with two hypothetical dilemmas involving trade-offs between different forms of the self (materials adapted from www.philosophersnet.com/games/). Dilemma 1 described a scenario in which participants could choose to travel to Mars by spaceship, or to be replicated by a transporter that would destroy their body but recreate an exact copy of it on Mars (see Supplementary materials). Dilemma 2 described a scenario in which participants were asked to imagine they had a fatal illness, and that although scientists were working on a cure, it
would not be ready in time to save them. Participants were asked to decide to either freeze their body to be later thawed and cured when the cure was discovered (however, the freezing process would likely kill their soul), or to let their body die from the disease but preserve their soul. Participants responded on a six-point scale, with endpoints: 1 = I would strongly prefer to let my body die, 6 = I would strongly prefer to be frozen. While both Dilemmas 1 and 2 measure participants’ relative value of their physical body, only Dilemma 2 contrasts the physical body with the soul. Responses to Dilemma 2 serve as our dependent variable because they indicate whether participants value their soul more or less than their physical body.

2.1.2. Other measures
At the end of the study, participants completed an 11-item Dualism scale comprised of questions on belief in soul, separation between mind and body, and materialism of the mind (e.g., “The true self is not governed by the brain, but by a person’s soul”), endpoints: 1 = Strongly Disagree; 7 = Strongly Agree), and a three-item religiosity index that measured frequency of religious attendance (1 = never, 5 = frequently), religiosity (1 = not at all, 5 = Strongly), and belief in God (1 = not at all, 5 = Strongly). Participants then reported the number of neuroscience and psychology courses previously taken, and completed demographic items on gender, age, ethnicity, and also and political ideology (1 = Strongly liberal, 5 = Strongly conservative). Finally, participants were probed for suspicion and debriefed.

2.2. Results

2.2.1. Manipulation check
No participants correctly guessed the hypothesis of the study. We compared the course titles selected for the Psychology and Neuroscience courses to test whether the course descriptions effectively manipulated participants’ perceptions of the mechanistic nature of these phenomena. Title selections for each course were dummy-coded as 1 or 2, with 1 reflecting the mechanistic title, and 2 reflecting the explanatory gap title. Choice of the mechanistic-explanation vs. explanatory-gap title was correlated with the Love and Morality courses, \( r(101) = .30, p = .002 \), and the mean of the two titles was computed. An independent sample t-test confirmed that subjects were more likely to select the mechanistic course titles in the neuroscience condition relative to the psychology condition, \( t(99) = 5.09, p < .001 \). Participants who answered both manipulation checks incorrectly were removed from further analyses, leaving \( n = 134 \) in the total sample including participants in the control condition.

2.2.2. Body–soul dilemmas
One-way ANOVAs on condition were conducted on both hypothetical dilemmas. We did not expect any effect on Dilemma 1 (teleportation), and no effect was observed for the omnibus ANOVA, \( F(2, 131) = 1.96, p = .14 \). But consistent with predictions, the omnibus ANOVA was significant for Dilemma 2 (body–soul tradeoff), \( F(2, 131) = 3.47, p = .03, \eta^2 = .05 \), and on the planned linear comparison (weights: neuroscience = −1, control = 0, psychology = 1) was also significant \( F(1, 131) = 7.11, p < .01 \). Participants in the neuroscience condition showed strongest preference to freeze their body and let their soul die \( (M = 3.59, SD = 1.77) \) compared to the control condition \( (M = 3.34, SD = 1.69) \), and participants in the psychology condition reported lowest preference \( (M = 2.57; SD = 1.74) \) to freeze their body and let their soul die.

2.2.3. Other measures
Reliability on the three religiosity items (attendance, religiousness, belief) was strong, (Cronbach’s \( x = .85 \)), we therefore created a religiosity index using the mean of the three items. The 11-item Dualism scale showed good scale reliability (Cronbach’s \( x = .84 \)). We calculated the correlations between Dualism scales, religiosity, Dilemma 1 (teleportation) and Dilemma 2 (freeze body/kill soul). The religiosity index was positively correlated with Dualism \( (r = .50, p < .001) \), and negatively correlated with the choice to freeze body/kill soul in Dilemma 2 \( (r = −.36, p < .001) \). Belief in dualism was also negatively associated with the choice to freeze body/kill soul \( (r = −.49, p < .001) \). None of the variables were correlated with Dilemma 1 (teleportation) \( (all \ p > .23) \), suggesting that value of the physical self measured by this dilemma was conceptually distinct from mind–body duality and other spiritual/metaphysical beliefs. One-way ANOVA on condition was not significant on either the Dualism scale \( (F < 1, p = .43) \) or on the religiosity index \( (F < 1, p = .66) \).

2.2.4. Discussion
In Experiment 1, belief in the soul was measured by a hypothetical tradeoff between body and soul. Reading psychological descriptions about love and morality accompanied by neuroscience explanations increased preference for body-continuity over the soul. Conversely, reading the psychological descriptions without neuroscience explanations (i.e., raising questions about the phenomenology without providing answers) increased preference for soul-continuity over the body. Manipulations did not affect preferences in a teleportation dilemma, suggesting the effect is not merely due to increased value for the physical body. Rather, the findings support our argument that belief in the soul and belief in mechanistic explanations of experience are reflexive explanations for understanding the psychological experience of the mind.

3. Experiment 2
In Experiment 1, exposure to a neuroscience course description increased relative value of the physical body over the soul, whereas reading a psychology course description on the same topics increased value of the soul over the physical body. Experiment 2 extended the findings of Experiment 1 to a behavioral domain. Drawing from popular culture, we examined participants’ willingness to “sell” their soul after being exposed to neuroscience information. In addition, Experiment 2 manipulated the explanatory strength of the neuroscience information (strong or weak explanations). This allowed us to test whether the
mere presence of neuroscience information would decrease belief in soul, or whether the neuroscience information must serve as a strong explanation for psychological phenomena while weak neuroscience information (i.e., an explanatory gap) may enhance lay belief in soul.

3.1. Method

Seventy-five undergraduates (37 women) volunteered to participate for partial course credit. Participants were assigned to one of three conditions: strong neuroscience, weak neuroscience, or a control condition. In both neuroscience conditions, participants read brief descriptions of neuroscience research on the experience of free will, love, and moral judgment. In the strong neuroscience condition, each description concluded with the statement that the phenomenon was completely explained by neural activity. In the weak neuroscience condition, each description concluded with a statement that the research could not explain the source of the phenomenon or how brain activity translates into the phenomenal experience (see all neuroscience descriptions in Appendix B). Following each passage, participants were asked to select one of two titles that best described the passage. In the control condition there was no passage.

3.1.1. Soul cards

Previous research has shown that most undergraduate participants refuse to sell their soul as part of a psychology experiment even if they do not believe in souls (Haidt & Bjorklund, 2000). Instead of buying participants’ souls, we made the experimental paradigm less threatening to subjects by using a novelty card to represent the soul. Following the course description manipulation, participants were brought into a separate room and were seated at a desk across from the experimenter. Each participant received a novelty soul card as a free gift for participating in the experiment. The front of the card was labeled “Soul ID card”, with a 35 digit ID code and a picture of a human form. The back of the card had spaces that read: “initials of carrier”, and “date of birth” for the cardholder to complete. The experimenter placed the soul card face down on the table and asked participants to write their date of birth and initials on the card. The card was turned over and participants were told it represented their soul and were given a moment to examine the card.

3.1.2. Bargaining

Participants were told that the card was theirs to keep if they wished or that they could try to sell it back to the experimenter for up to $10. The rules of the bargaining procedure were then explained as follows: If participants chose to sell their soul cards, they would write down a minimum asking price (up to $10) on a piece of paper, that represented the smallest amount of money for which they would be willing to sell the card. After writing down their price, participants randomly drew a dollar amount between $1 and $10 (written on slips of paper) from a box. If the amount randomly drawn was equal to or greater than the participant’s asking price, the participant would have to sell the soul card in exchange for the amount drawn from the box. However, if the participant drew a number that was less than the asking price, the participant would keep the soul card and would not get any money. Thus, participants should be motivated to generate a minimum price for the card that reflects its subjective value. Once the bargaining procedures were clear, participants decided to either keep or try to sell their soul card in the bargaining game. Those participants who chose to keep their card (i.e., not play the bargaining game), were then asked for a hypothetical minimum price (above $10, no limit), that they would accept to sell the card. Special care was taken during debriefing to reassure subjects that the card was not a true embodiment of their soul, and that no real purchase of their soul had transpired.

3.1.3. Other measures

After the bargaining procedure, participants completed a 13-item Dualism scale (see Supplementary materials). Participants reported their current college major and number of science classes previously taken. Finally, participants completed demographic items on gender, age, ethnicity, and also frequency of religious attendance (1 = never, 5 = frequently), religiosity (1 = not at all, 5 = strongly), and belief in God (1 = not at all, 5 = strongly).

3.2. Results

3.2.1. Manipulation check

Choices of the passage title (mechanistic-explanation vs. explanatory-gap titles) in the Psychology/Neuroscience conditions were summed across the three passages, and an independent sample t-test confirmed that subjects were more likely to select the mechanistic titles in the Strong vs. Weak neuroscience condition, t(50) = 5.06, p < .001. All participants answered at least one manipulation check item correctly and were retained in the subsequent analyses.

3.2.2. Bargaining

Participants who played the bargaining game were coded as willing to sell the soul card, regardless of whether the bargain resulted in a successful sale of the card. The proportion of participants who chose to sell their soul card in each condition was analyzed by chi-square (see Table 1). As predicted, the linear by linear association on condition was significant, χ² (1, N = 75) = 5.02, p = .03. Relative to the control condition (52% sell soul card), participants exposed to strong neuroscience explanations were more likely to sell their soul card (72%), whereas participants exposed to weak neuroscience explanations were less likely to sell their soul card (40%).

We also examined the reported minimum prices to sell the card. For those who opted to engage in the bargaining game, prices ranged from 1 to 10 dollars. But those who did not play the game reported a hypothetical minimum price that had no upper limit. The hypothetical minimum prices were not recorded for eight participants who chose to keep the card. For the remaining subjects who kept the card, minimum prices ranged between one thousand dollars and “priceless” (see Fig. 1). The skewed variance and limitless upper range of this data made it unfeasible to...
analyze the differences statistically, but simultaneously reveals the powerful difference between those who were willing to sell the card vs. those who were not. Subjects who chose not to play the bargaining game valued the soul card considerably more than those who chose play the bargaining game, suggesting it took on some important value beyond just a piece of paper.

3.2.3. Other measures

The three religious items (attendance, religiousness, belief) showed good scale reliability, Cronbach’s α = .90. We therefore computed the mean of these items to create a religiosity index. The Dualism scale also showed good reliability (Cronbach’s α = .74), and was positively correlated with the religiosity index (r = .49, p < .001). One-way ANOVA on condition was not significant on either the Dualism scale (F(2, 72) = 1.84, p = .17) or on the religiosity index (F < 1, p = .57). However, willingness to sell the soul card was negatively correlated with both the religiosity index (r = −.25, p = .01), and the Dualism scale (r = −.29, p = .01).

3.3. Discussion

In Experiment 2, willingness to sell a soul card varied as a function of exposure to strong versus weak neuroscience explanations for psychological phenomena. Exposure to strong neuroscience explanations (for the experience of love, moral judgment, and conscious will) increased participants’ willingness to sell their soul card. This neuroscience information provided an alternative material explanation for important psychological phenomena, suggesting an explanatory competition between the soul and brain as different explanations of mind. On the other hand, weak neuroscience information decreased willingness to sell the soul card. This suggests that mere awareness of neuroscience does not diminish belief in the soul, rather it is the apparent power of neuroscience as an explanation for psychological phenomena. Indeed, weak neuroscience explanations further diminished willingness to sell the soul card, evidence that an apparent explanatory gap can enhance metaphysical explanations as an alternative means of understanding the experience of mind.

It is also worth noting that in both Experiments 1 and 2, manipulations did not impact responses in a self-reported Dualism scale, but the scale was correlated with religiosity and the key dependent measures. It is possible that the Dualism scale was too far removed from the manipulation to be affected, but another possibility is that the Dualism scale has better validity as a measure of individual differences in belief (or “trait” beliefs), rather than temporary changes in belief, (or “state” beliefs). People can be reluctant to alter their explicit statements about religious and spiritual belief after a manipulation of this kind, but behavioral measures of belief are more sensitive to gut-level feelings that can be affected by explanatory information. Unlike the explicit statements in the Dualism scale, the key dependent measures in both studies (the hypothetical dilemma and selling the soul-card) used a symbolic action on the soul. This may be essential to their effectiveness, because these measures allow psychological leeway to alternatively

![Fig. 1. Minimum selling prices for the Soul Card, by subjects who opted to sell or keep in the bargaining game.](image)
treat the soul as either real or fiction, without violating explicit religious convictions.

4. Conclusions

People have long believed in the soul, in part as a way to understand the source of the mind and consciousness. But recently neuroscience research has made important discoveries about the brain’s involvement in essential psychological experiences that may impinge on these long-held beliefs. The present research found that belief in the soul was inversely related to the availability of neuroscience explanations of the mind. Importantly, the direction of change depended on the perceived explanatory power of the neuroscience information: whereas strong neuroscience explanations decreased belief in the soul, weak neuroscience explanations increased belief in the soul. Consistent with a causal discounting account, these findings suggest that acceptance of physical vs. metaphysical kinds of explanation are inversely related, and may be used reflexively as a means to understand the mind.

Although these studies demonstrated a hydraulic relation between neuroscience evidence and belief in soul, we should be clear that people do not necessarily choose one explanation over another. Indeed, in the absence of strong evidence for one kind of explanation, it may be most common for people to seek a synthesis between physical and metaphysical folk theories and use aspects of each to satisfy different explanatory needs (Legare, Evans, Rosenberg, & Harris, 2012). Descartes himself argued that the pineal gland was the “seat of the soul” – a physical structure where the soul interacted with the body (Descartes, 1641). Today many people subscribe to a kind of modern dualism that accepts the brain as responsible for computational cognitive functions (i.e., ability to think), but reserves the soul as an explanation for personal aspects of mind, such as emotions (Richert & Harris, 2008). As the public love affair with neuroscience continues, the feeling of a personal essence of self that we all experience will certainly remain. However, the present results do illustrate that relative belief in physical versus metaphysical descriptions of mind change in response to new neuroscience information, and the trajectory of neuroscience research may forecast a shift in the relative emphasis on the soul as an explanatory tool.

Finally, these findings have some interesting implications for the study of neuroscience itself. It may be tempting to conclude that neuroscience research will replace belief in the soul as it uncovers more information on the brain’s role in our most essential human experiences. Important however, neuroscience research still does not seem to fully capture the subjective experience of consciousness that compels us to “feel” a mind separate from our bodies (the root of the hard problem of consciousness). Indeed, the present findings also imply that limitations of neuroscience provide explanatory space for the soul to flourish in the age of fMRI research. There are also other reasons and motivations to believe in the soul (e.g., fear of death), and potential social costs to losing belief in the soul (e.g., decreased moral responsibility; Vohs & Schooler, 2008). It is therefore possible that advances in neuroscience could create a backlash for threatening the valued belief in the soul. What then, should be neuroscientists’ responsibility in framing their results? To protect the cherished beliefs of others by diminishing the explanatory power of neuroscience findings? Or to take a confrontational stance against the soul by framing results as powerful explanations, and so dissuade the public from the perils of magical thinking? Science’s responsibility must be only to the truth, of course, but given the serious issues of both underselling and overclaiming the implications of neuroscience research, the need for clear and unbiased representation of results is ever more important.

Acknowledgments

We thank Evan Branson, Da Yae Keum, Claire Houtsma, Regina Mendez, Tessa Rainey, Agnes Strojewska, and Pin-ya Tseng for assistance in data collection. This publication was made possible from grant support from the John Templeton Foundation, awarded to Jesse Preston. The opinions expressed in this publication are those of the authors and do not necessarily reflect those of the John Templeton Foundation.

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.cognition.2012.12.003.

References