Free will and punishment: A mechanistic view of human nature reduces retribution

Shariff, A.F.; Greene, J.D.; Karremans, J.C.T.M.; Luguri, J.B.; Clark, C.J.; Schooler, J.W.; Baumeister, R.F.; Vohs, K.D.

2014, Article / Letter to editor (Psychological Science, 25, 8, (2014), pp. 1563-1570)

Doi link to publisher: https://doi.org/10.1177/0956797614534693

Corrigendum

Psychological Science
1
© The Author(s) 2017
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797617750381
www.psychologicalscience.org/PS

Corrigendum: Free Will and Punishment: A Mechanistic View of Human Nature Reduces Retribution

Original article: Shariff, A. F., Greene, J. D., Karremans, J. C., Luguri, J. B., Clark, C. J., Schooler, J. W., . . . Vohs, K. D. (2014). Free will and punishment: A mechanistic view of human nature reduces retribution. *Psychological Science, 25*, 1563–1570. doi:10.1177/0956797614534693

We are very grateful to Nick Brown and James Heathers for bringing the following errors to our attention.

Corrected Mediation Figure for Study 3

In Figure 2, we showed how blameworthiness mediated the effect that exposure to neuroscientific research had on punishment. In the original figure, we reported a regression coefficient of −0.50 for the *b* path (between blameworthiness and punishment). This should have been +0.50, indicating a positive relationship, and the figure is now being changed accordingly. The positive relationship (i.e., higher blame being related to higher punishment) was the predicted one and the one reported in the main text. In addition, the coefficients were reported as standardized when they were, in fact, unstandardized; the caption and figure are now being updated to reflect these changes.

Corrected Labeling of Manipulation Check in Study 3

In the Method for Study 3, we erroneously reported that we used the Free Will and Determinism Plus (FAD+) scale from Paulhus and Carey (2011). In reality, we used an older version of the scale (the FAD; Paulhus & Margesson, 1994), and the article will be changed to reflect this.

References

Paulhus, D. L., & Carey, J. M. (2011). The FAD-Plus: Measuring lay beliefs regarding free will and related constructs. *Journal of Personality Assessment, 93*, 96–104.

Paulhus, D. L., & Margesson, A. (1994). *Free Will and Determinism (FAD) scale*. Unpublished manuscript, Department of Psychology, University of British Columbia, Vancouver, Canada.

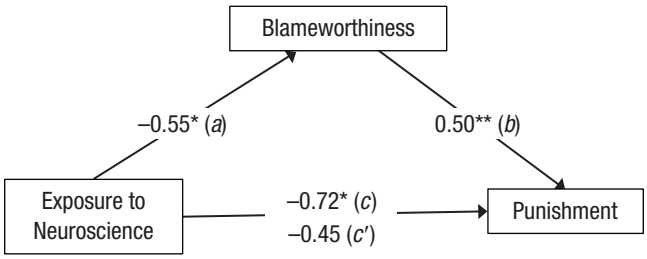

Fig. 2. Effect of exposure to neuroscientific research on punishment recommendations as mediated by perceived blameworthiness of the offender in Study 3. Unstandardized regression coefficients are shown along the paths; *c*′ represents the mediated effect of condition on punishment recommendations, taking into account the indirect effect of blameworthiness. Asterisks indicate significant effects (*$p < .05$, **$p < .001$).

Research Article

Psychological Science
2014, Vol. 25(8) 1563–1570

DOI: 10.1177/0956797614534693
pss.sagepub.com

Free Will and Punishment: A Mechanistic View of Human Nature Reduces Retribution

Azim F. Shariff[1], Joshua D. Greene[2], Johan C. Karremans[3], Jamie B. Luguri[4], Cory J. Clark[5], Jonathan W. Schooler[6], Roy F. Baumeister[7], and Kathleen D. Vohs[8]

[1]Department of Psychology, University of Oregon; [2]Department of Psychology, Harvard University; [3]Department of Social and Cultural Psychology, Behavioral Science Institute, Radboud University Nijmegen; [4]Department of Psychology, Yale University; [5]School of Social Ecology, University of California, Irvine; [6]Department of Psychology, University of California, Santa Barbara; [7]Department of Psychology, Florida State University; and [8]Carlson School of Management, University of Minnesota

Abstract

If free-will beliefs support attributions of moral responsibility, then reducing these beliefs should make people less retributive in their attitudes about punishment. Four studies tested this prediction using both measured and manipulated free-will beliefs. Study 1 found that people with weaker free-will beliefs endorsed less retributive, but not consequentialist, attitudes regarding punishment of criminals. Subsequent studies showed that learning about the neural bases of human behavior, through either lab-based manipulations or attendance at an undergraduate neuroscience course, reduced people's support for retributive punishment (Studies 2–4). These results illustrate that exposure to debates about free will and to scientific research on the neural basis of behavior may have consequences for attributions of moral responsibility.

Keywords

free will, punishment, morality, responsibility, blame, open materials

Received 3/14/13; Revision accepted 4/14/14

Most people believe that humans have free will (Nahmias, Morris, Nadelhoffer, & Turner, 2005). However, long-standing controversies remain as to what forms of free will can actually exist alongside the known laws of nature (see Baer, Kaufman, & Baumeister, 2008). Although few people deny that humans regularly make uncoerced choices and exercise self-control, many scientists and philosophers have taken issue with the idea that conscious humans can generate spontaneous choices and actions not fully determined by prior events (e.g., Bargh, 2008). According to this challenge, free will of this kind appears to be incompatible with a scientific understanding of the world as a mechanical system in which all events are fully determined by prior or random events. Many researchers have argued that this form of free will is an illusion that grows less believable as research supporting the mechanistic causes of human behavior accumulates (Greene & Cohen, 2004; Wegner, 2002). As anti-free-will viewpoints reverberate beyond academia, and legal arguments and popular-press articles point to ever-more mechanical causes for human behavior (Greene & Cohen, 2004; Nahmias, 2011; Wolfe, 1997), questions about the psychological and societal consequences of reduced free-will beliefs have emerged (Schooler, 2010; Shariff, Schooler, & Vohs, 2008).

Free-will beliefs underlie perceptions of moral responsibility (Eshleman, 2004; Nahmias et al., 2005). Legal and moral decisions often rest on whether one *should* have

Corresponding Author:
Azim F. Shariff, 1227 University of Oregon, Eugene, OR 97401
E-mail: shariff@uoregon.edu

acted differently, which presupposes that one *could* have acted differently. When genuine choice is deemed impossible, condemnation is less justified. Purely natural phenomena, such as viruses and hailstorms, are not held morally responsible for the damage they cause because they are not perceived as freely choosing their actions. The rejection of free will for humans could similarly undermine attributions of responsibility, both for oneself and for others, rendering human actions akin to other natural phenomena.

As a consequence, diminishing people's belief in free will may likewise weaken their belief in moral responsibility, and potentially license them to transgress. Indeed, prior research found that participants whose free-will beliefs were experimentally diminished were less helpful and more likely to lie, cheat, steal, and act aggressively than were participants in control conditions (Baumeister, Masicampo, & DeWall, 2009; Vohs & Schooler, 2008). Such antisocial outcomes lend support to concerns about the negative social consequences that could follow a broad erosion of free-will beliefs.

In the current work, we considered the opposite side of that coin. We tested whether reduced belief in free will would lead people to see others' bad behavior as less morally reprehensible, resulting in less retributive punishment.

Free Will, Retributivism, and Consequentialism

Humans respond to transgressions with an urge to exact punitive costs on the transgressor (e.g., Buckholtz et al., 2008; Nelissen & Zeelenberg, 2009; Smith, 1759). One theory of punishment, labeled retributivism, holds that punitive urges reflect normative moral principles based on universal norms of reciprocity and fairness: People who harm others should be harmed themselves. In this view, reciprocity is moral justification for punishment, independent of any other benefits the punishment may bring. Retributivism, in other words, holds that the point of punishment is to extract suffering from the transgressor as "just deserts." Retributivism is contrasted with consequentialist justifications for punishment, which depend on the utilitarian benefits of punishment (e.g., rehabilitation or deterrence). According to consequentialism, selection of punishment should be based on what has the best social consequences, regardless of how much or how little the transgressor suffers (if at all), or indeed whether the transgressor even deserves punishment.

Research has provided some evidence for a correlation between beliefs about punishment and beliefs in free will. For instance, Krueger, Hoffman, Walter, and Grafman (in press) found that people with greater belief in free will also tended to be more punitive, though only for transgressions that were not especially emotionally arousing. Theorists have argued, however, that only retributivist, and not consequentialist, motivations for punishment should depend on the strong and embattled form of free will discussed in our opening paragraph (Greene & Cohen, 2004). Because consequentialism is unconcerned with whether someone deserves punishment in an ultimate sense, and is instead focused on whether punishment would have a positive utilitarian effect, it is not threatened by mechanistic arguments against free will. In contrast, retributivism is based on transgressors being blameworthy and deserving suffering because of their transgressions, so anything that diminishes that sense of deservingness—such as an inability to have freely chosen not to commit the transgression—diminishes the justification for retributive punishment. Though these are theoretical connections, there is evidence to suggest that people do base their moral decision making, in part, on their philosophical worldview (e.g., Rai & Holyoak, 2013; Vohs & Schooler, 2008). Therefore, we predicted that people's free-will beliefs should be related primarily to their support for retributive punishment, and further, that these beliefs are causal factors. Diminished beliefs in free will should diminish blame, which should, in turn, diminish the endorsement of retributive punishment.

The Present Research

In Study 1, we tested the relationship between free-will beliefs and attitudes toward consequentialist and retributivist punishment. We then moved to experimental methods, testing whether punitive attitudes were affected by diverse manipulations aimed at diminishing free-will beliefs. In Study 2, we had participants read about challenges to free will, whereas in Studies 3 and 4, participants were exposed to research about the neural mechanisms underlying human action. The outcome measures in these experiments were retributivist attitudes toward criminality. We predicted that reduced belief in free will would predict weaker retributivist punishment.

Study 1: Correlations Between Free-Will Beliefs and Retribution

Study 1 tested how free-will beliefs relate to attitudes about both retributive and consequentialist punishment. Given that retributive punishment springs from moral blame (Greene & Cohen, 2004) whereas consequentialist punishment is intended to benefit society, we predicted that support for retributive punishment would be positively related to free-will beliefs, whereas support for consequentialist punishment would be unrelated.

Table 1. Results of Regression Analyses Predicting Support for Retributive Punishment and Consequentialist Punishment in Study 1

Predictor	Retributive punishment	Consequentialist punishment
Free-will belief (no controls)	0.242**	–0.024
Free-will belief (controlling for the variables below)	0.164*	0.018
Age	0.088	0.156*
Gender (male)	0.050	0.045
Education	–0.139*	0.056
Religiosity	–0.158*	0.034
Political ideology (social)	0.323**	–0.215*
Political ideology (economic)	0.003*	–0.013

Note: The table reports standardized regression coefficients. Higher values for political ideology indicate a more conservative position.
*$p < .05$. **$p < .001$.

Method

Two hundred forty-four Americans (147 female; mean age = 36.81 years) participated online via Amazon's Mechanical Turk. Respondents completed the seven-item Free Will subscale of the Free Will and Determinism Plus scale (FAD+; Paulhus & Carey, 2011; $\alpha = .88$. Items (e.g., "People have complete control over the decisions they make") were rated from 1 (*strongly disagree*) to 5 (*strongly agree*). FAD+ scores have been shown to be moderately positively associated with religiosity, belief in a just world, and locus of control (Paulhus & Carey, 2011).

In order to measure attitudes toward retributivist and consequentialist motivations for punishment, we had participants read descriptions of retributivism and consequentialism as motivations for punishment and then indicate on two separate Likert scales (1 = *strongly disagree*, 7 = *strongly agree*) how important retributivism and consequentialism should be in determining motivations for criminal punishment. Participants also completed questions about their education, religiousness, political ideology (including separate ratings for economic and social issues), and demographics (for the materials used in Study 1, see Methodological Details and Materials in the Supplemental Material available online).

Results

Zero-order correlations indicated that the retributivism and consequentialism scales were moderately negatively correlated, $r(243) = -.36$, $p < .001$.

In order to test for relationships with free-will beliefs, we regressed retributivism and consequentialism scores on Free Will subscale scores. As predicted, stronger belief in free will predicted greater support for retributive punishment, $\beta = 0.24$, $p < .001$, but was not predictive of support for consequentialist punishment, $\beta = -0.02$, $p = .72$. Effects remained significant when we statistically controlled for age, gender, education, religiosity, and economic and social political ideology (Table 1).

Study 1 supported the hypothesis that free-will beliefs positively predict punitive attitudes, and in particular, retributivist attitudes. Though we controlled for potential third-variable explanations, such as political ideology, the correlational design did not allow us to determine whether a reduction of free-will beliefs would lead to a resultant reduction in retributivism. Therefore, in Studies 2 through 4, we moved to experimental methods to test the causal relationship between free-will beliefs and support for retributive punishment by manipulating free-will beliefs directly.

Study 2: Manipulated Free-Will Beliefs and Retributivism

Method

Participants. Forty-six students (30 female; mean age = 20.44 years) participated for partial course credit.

Procedure. Told that they were taking part in unrelated experiments, participants first completed a free-will manipulation. They were randomly assigned to read one of two passages from Crick's (1994) *The Astonishing Hypothesis*. In the anti-free-will condition, the passage rejected free will and advocated a mechanistic view of human behavior. In the neutral condition, the passage was unrelated to free will. This task has been previously validated by Vohs and Schooler (2008), and testing the manipulation in an independent sample revealed that the passages led to the expected differences in free-will beliefs, as measured by a single item (see Methodological Details and Materials in the Supplemental Material), $t(205) = 2.55$, $p = .01$.

Next, participants read a fictional vignette involving an offender who beat a man to death. Acting as hypothetical jurors, participants recommended the length of the prison sentence (if any) that this offender should serve following a 2-year, nearly 100%-effective, rehabilitation treatment. The notion that the offender had been rehabilitated

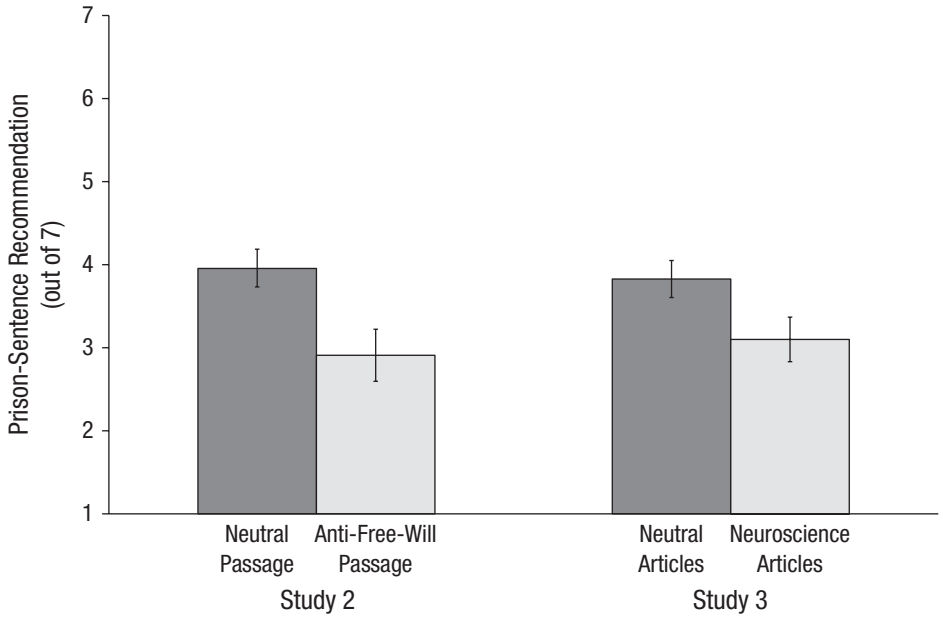

Fig. 1. Mean recommended prison sentence for a hypothetical criminal in Study 2 (neutral and anti-free-will conditions) and Study 3 (neutral and neuroscience conditions). Error bars represent ±1 *SEM*.

was used in order to isolate participants' desire for punishment as retribution.[1] The passage further focused participants on retributive, rather than consequentialist, punishment by noting that the prosecution and defense had agreed that the rehabilitation would prevent recidivism and that any further detention after rehabilitation would offer no addition deterrence of other potential criminals.

Participants then chose among seven punishment options: treatment time only, with no imprisonment (1); 2 years of imprisonment posttreatment (2); 5 years of imprisonment posttreatment (3); 10 years of imprisonment posttreatment (4); 25 years of imprisonment posttreatment, with the chance of parole after 15 years (5); 25 years of imprisonment posttreatment, with no chance of parole (6); and life imprisonment, with no chance of parole (7). Finally, participants completed a demographics questionnaire and suspicion probe (for the materials used in this experiment, see Methodological Details and Materials in the Supplemental Material).

Results

As predicted, participants who read the anti-free-will passage recommended significantly lighter prison sentences than participants who read the neutral passage ($M = 2.91$, $SD = 1.08$, vs. $M = 3.96$, $SD = 1.49$), $t(44) = 2.71$, $p < .05$, Cohen's $d = 0.82$ (Fig. 1). Specifically, participants whose free-will beliefs had been experimentally diminished recommended roughly half the length of imprisonment (~5 years) compared with participants who read the neutral passage (~10 years).

Therefore, Study 2 demonstrated that experimentally diminishing free-will beliefs alters legal judgments by reducing inclinations for retributive punishment. In Study 3, we aimed to bolster this finding using a subtler manipulation and measuring ratings of blameworthiness in order to directly test the hypothesis that beliefs about free will undergird judgments of moral responsibility.

Study 3: Systematic Exposure to Neuroscience and Retributivism

In Study 2, we diminished participants' free-will beliefs by having them read a quite opinionated essay written by a Nobel laureate (Crick, 1994). In Study 3, we used a subtler manipulation that may represent how people's free-will beliefs are more commonly challenged; participants read articles made to look like they were popular-science magazine articles on findings from recent neuroscientific research. These articles did, in fact, describe actual neuroscience findings implying that human behavior is caused mechanistically. Crucially, the passages made no mention of free will, thereby allowing participants to draw their own conclusions about its relevance.

We hypothesized that relative to exposure to scientific views on other topics, exposure to research implying a mechanistic view of human action would reduce belief in free will, and thereby retributive tendencies in sentencing

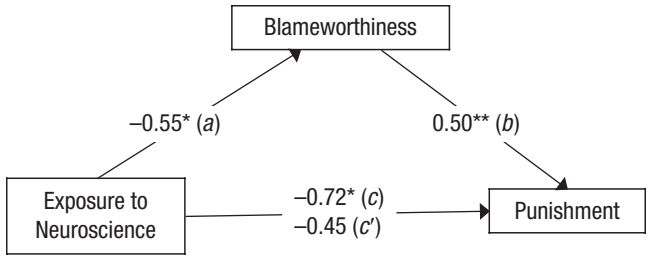

Fig. 2. Effect of exposure to neuroscientific research on punishment recommendations as mediated by perceived blameworthiness of the offender in Study 3. Unstandardized regression coefficients are shown along the paths; *c′* represents the mediated effect of condition on punishment recommendations, taking into account the indirect effect of blameworthiness. Asterisks indicate significant effects (*$p < .05$, **$p < .001$).

recommendations. Moreover, we predicted that exposure to mechanistic neuroscience would reduce the extent to which a transgressor was perceived as blameworthy, and that this assessment would mediate the effect of condition on sentencing recommendations.

Method

Participants. Ninety-one undergraduates participated for partial course credit. Four participants were excluded (3 for suspicion, 1 for admitted intoxication), which left a final sample of 88 participants (61 female; mean age = 20.81 years).

Procedure. Told that they were taking part in unrelated experiments, participants first read two popular-science articles made to look as if they were from the *Scientific American* and *New Scientist* Web sites. Participants in the neuroscience condition read articles on brain-imaging studies that showed dissociations between motor actions and people's perception of conscious intention (Osborne, 2003; Soon, Brass, Heinze, & Haynes, 2008). Participants in the neutral condition read articles on nuclear power and natural headache remedies. In neither condition did the material mention free will, morality, or responsibility. As in Study 2, pretesting with an independent sample revealed that the neuroscience passages led to lower free-will beliefs than did the neutral science passages, $t(198) = 2.35$, $p = .02$. This confirms that relative to other scientific descriptions, descriptions of scientific research promoting a mechanistic view of human behavior diminish belief in free will.

For the ostensible second study, participants read and responded to the vignette from Study 2. In addition to recommending a prison sentence, they rated the offender's blameworthiness (1 = *not at all deserving of blame*, 7 = *completely deserving of blame*). Participants then completed the Free Will and Determinism scale (FAD; Paulhus & Margesson, 1994) as a manipulation check and the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) so that we could account for mood effects that might emerge from challenges to intuitions about the mechanisms behind human action. Finally, participants completed a demographics questionnaire and suspicion probe. (For the materials used in this experiment, see Methodological Details and Materials in the Supplemental Material.)

Results

Mood. Neither positive nor negative mood differed significantly by condition (ps > .4). Higher levels of positive affect corresponded to lower attributions of blame, $r(86) = -.34$, $p = .002$. No other relationships involving mood were found.

Manipulation check. Results for the manipulation check were consistent with the independent test of the manipulation; participants who read the neuroscience articles reported marginally lower belief in free will compared with participants who read the neutral articles, $F(1, 87) = 3.46$, $p = .07$.

Punishment and blameworthiness. As predicted, participants who read the neuroscience articles recommended significantly shorter prison sentences than did participants who read the other science articles ($M = 3.10$, $SD = 1.48$, vs. $M = 3.83$, $SD = 1.77$), $t(86) = 2.09$, $p = .04$, $d = 0.45$ (Fig. 1) and blamed the transgressor less ($M = 5.48$, $SD = 1.17$, vs. $M = 6.03$, $SD = 0.92$), $t(86) = 2.40$, $p = .02$, $d = 0.52$. Recommended punishment and perceived blameworthiness were significantly correlated, $r(88) = .37$, $p < .001$. Controlling for mood did not markedly change the main results.[2]

Mediation. We predicted that reading the neuroscience articles would decrease participants' perception that the offender was morally blameworthy, and that this effect on perceived blameworthiness would account for the lighter sentencing recommendations in the neuroscience condition than in the neutral condition. Consistent with predictions, bootstrapping analyses (10,000 resamples; Preacher & Hayes, 2004) indicated that perceived blameworthiness mediated the effect of condition on sentencing recommendations (95% confidence interval = [−0.68, −0.03]; Fig. 2).

Study 3 bolstered the results of Study 2 in several ways. First, it replicated the finding that experimental manipulations that decrease free-will beliefs also reduce retributive punishment. Second, the mediational effect of perceived blameworthiness made a strong case for the role of moral responsibility in the effect of diminished free-will belief on retribution. Third, Study 3 used a

manipulation that did not mention free-will beliefs but instead influenced them by presenting participants with scientific findings suggesting that human behavior is mechanistic.

Study 4: Neuroscience Education and Retribution

In Study 4, we delved further into the issue of naturalistic manipulations of free-will beliefs. Instead of using a laboratory-based manipulation, we employed a naturalistic method by which people learn about mechanistic causes underlying human action: university neuroscience classes. Participants in an introductory cognitive-neuroscience class judged the appropriate punishment for a hypothetical criminal. They did so on both the first and last days of class, 10 weeks apart. We predicted that learning about the brain would reduce support for retributive punishments, a result that would conceptually replicate the prior studies. We further predicted that students' attitude change across time would correspond to measures of learning in the class. We also included students from a nonneuroscience class as a neutral comparison group.

Method

Participants. Students from an introductory cognitive-neuroscience class (n = 34) and an introductory geography class (n = 36) participated in exchange for the chance to win a raffle prize. Two students who were simultaneously taking both classes were dropped from analysis, which left 68 participants (39 female; mean age = 20.44 years).

Procedure. On the first day of class, students completed a shortened version of the punishment scenario from Studies 2 and 3. They were also asked about their perceived knowledge of the brain (relative to other students from the university), age, sex, and current classes. Ten weeks later, at the courses' final class, students completed another survey with the same measures, followed by a question about their anticipated class standing. Instructors for both classes were blind to the hypothesis. (For the materials used in this experiment, see Methodological Details and Materials in the Supplemental Material.)

Results

Supporting predictions, a paired-sample comparison showed that prison-sentence recommendations decreased from the start to the end of the neuroscience class, from 3.41 to 2.91 on the 7-point scale, $t(33) = 2.15$, $p = .04$, $d = 0.44$. No change was found for students in the geography class (Ms = 3.32 on the first day vs. 3.08 on the last day), $t(33) = 0.94$, $p = .41$, though the spreading interaction between the classes did not reach statistical significance, $F(1, 68) = 0.22$. $p = .64$.

Were changes in retributivism related to students' learning outcomes? In neither class did grades predict changes in sentence recommendations. However, for students in the neuroscience course, decrease in the length of sentence recommendations was strongly correlated with increase in students' self-reported knowledge of the brain from the first to the last class ($r = -.45$, $p = .01$). In the geography course, there was no correlation between self-reported changes in brain knowledge and sentencing recommendations ($r = .02$, $p = .90$). The strength of this relationship differed significantly between the classes (Steiger's $Z = 2.00$, $p = .05$), which suggests that the decrease in punitiveness directly corresponded to what students believed they had learned in the neuroscience class.

General Discussion

Convergent results across a correlational study and three experiments with diverse manipulations consistently demonstrated that shifting from a belief in free will toward a mechanistic view of human behavior reduces support for retributive punishment. Study 1 found that individual differences in free-will beliefs predict retributive, but not consequentialist, motivations for punishment. Study 2 found that experimentally diminishing free-will beliefs through anti-free-will arguments diminished retributive punishment, suggesting a causal relationship. Studies 3 and 4 found that exposure to neuroscience implying a mechanistic basis for human action—either reading popular-science articles or taking an introductory neuroscience class in college—similarly produced a reduction in retributivism. (For a summary table of all results, as well as results of replication attempts not mentioned here, see Methodological Details and Materials in the Supplemental Material.) These results suggest that shifts in people's philosophical worldview about free-will beliefs, even through simply learning about the brain, can affect people's attitudes about moral responsibility, with potentially broad social consequences.

Retributivism plays an important role in the justice system. Historically, much of the motivation for legal punishment has been an institutionalized attempt to sate the public's retributive desires (Smith, 1759). Legal historian Stephen (1883) famously wrote that "the sentence of the law is to the moral sentiment of the public what a seal is to hot wax" (p. 423). In recent years, justice researchers and advocates have argued for a switch from retributive to *restorative* justice—a consequentialist approach aimed at repairing the moral imbalances caused by transgressions (Braithwaite, 2002). The current findings suggest that changing attitudes about free will and responsibility may be important to this evolution of legal thinking.

That mere exposure to modern neuroscience can be sufficient to reduce retributivist motivations may be particularly relevant to court cases. The explicit existence of free will may be rarely debated in court, but neuroscientific evidence often is. Indeed, recent research showed that judges afforded shorter sentences to hypothetical psychopathic criminals when the description of the criminals' psychopathy included a biomechanical component, compared with when it did not (Aspinwall, Brown, & Tabery, 2012). Our findings likewise suggest that merely presenting such a perspective may move judges and jurors toward being less punitive and less retributive in general.

Whereas previous research showed that diminished beliefs in free will encourage antisocial, immoral behavior, the current findings expand this story. One explanation for the prior findings is that participants may have used the anti-free-will arguments as an excuse for moral laxity—taking advantage of apparently scientifically valid justifications in order to abandon self-control (Baumeister et al., 2009; Vohs & Schooler, 2008). In contrast, our studies offered no immediate benefit to participants for being more punitive. Although our data do not negate the idea that lowered free-will beliefs provide an excuse for self-interested behavior, they suggest that diminished free-will beliefs are more than excuses for selfishness: They appear to provoke a genuine decline in belief in moral responsibility. That such changes were shown to occur over the duration of a university neuroscience course (Study 4) suggests that as more people inside and outside of academia learn about neuroscience, moral attitudes may shift in response.

Whether these shifts are desirable is open to debate. Clearly, punishment serves important functions. Indeed, the presence of norms ensuring the punishment of transgressors is essential for group cohesion (Fehr & Gächter, 2002). Free-will beliefs may, therefore, serve an important cultural function in both encouraging the feelings of responsibility that motivate people to behave ethically and producing the moral outrage and retributive desires necessary to motivate costly but necessary punishment.

On the other hand, although diminished free-will beliefs reduce retributivist motivations, Study 1 suggests that the motivation to punish in order to benefit society (consequentialist punishment) may remain intact, even while the need for blame and desire for retribution are forgone. Thus, a societal shift away from endorsing free will could occur without disrupting the functional role of punishment. Society could fulfill its practical need for law and order, leaving the social benefits of punishment intact while avoiding the unnecessary human suffering and economic costs of punishment often associated with retributivism (Greene & Cohen, 2004; Tonry, 2004).

What is clear is that the belief in free will is intertwined with moral, legal, and interpersonal processes. As the mechanistic worldview espoused by many scientists, and particularly psychologists, gains attention (e.g., Gazzaniga, 2011; Monterosso & Schwartz, 2012; Nichols, 2011), the impact of these trends—good, bad, or both—calls for understanding.

Author Contributions

A. F. Shariff, J. D. Greene, J. C. Karremans, J. W. Schooler, and K. D. Vohs developed the study concept. A. F. Shariff and J. B. Luguri conducted the studies. A. F. Shariff analyzed and interpreted the data. A. F. Shariff drafted the manuscript, and C. J. Clark, J. D. Greene, J. C. Karremans, J. B. Luguri, J. W. Schooler, K. D. Vohs, and R. F. Baumeister provided critical revisions. All authors approved the final version of the manuscript for submission.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This project was supported in part by the John D. and Catherine T. MacArthur Foundation (Award 07-89249-000-HCD), by the Regents of the University of California, and by the John Templeton Foundation. The content of this publication does not necessarily reflect the views of the funders.

Supplemental Material

Additional supporting information may be found at http://pss.sagepub.com/content/by/supplemental-data

Open Practices

All materials have been made publicly available via Open Science Framework and can be accessed at osf.io/dy3pm. The complete Open Practices Disclosure for this article can be found at http://pss.sagepub.com/content/by/supplemental-data. This article has received the badge for Open Materials. More information about the Open Practices badges can be found at https://osf.io/tvyxz/wiki/view/ and http://pss.sagepub.com/content/25/1/3.full.

Notes

1. Though 100%-effective treatments do not yet exist, no participant noted explicit doubts regarding the effectiveness of the treatment. Although including this treatment in the vignette was necessary to circumvent rehabilitative motivations in this study, future research might include a more generalizable dependent measure.

2. Analyses controlling for PANAS scores confirmed that observed effects were due to mood. The effects of condition on FAD score, punishment, and blameworthiness in these analyses remained significant, $F(1, 87) = 5.98$, $p = .02$; $F(1, 84) = 5.06$, $p = .03$; and $F(1, 87) = 5.77$, $p = .04$, respectively.

References

Aspinwall, L. G., Brown, T. R., & Tabery, J. (2012). The double-edged sword: Does biomechanism increase or decrease judges' sentencing of psychopaths? *Science*, *337*, 846–849.

Baer, J., Kaufman, J. C., & Baumeister, R. F. (Eds.). (2008). *Are we free? Psychology and free will*. Oxford, England: Oxford University Press.

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

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.

Braithwaite, J. (2002). *Restorative justice & responsive regulation*. Oxford, England: Oxford University Press.

Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., & Marois, R. (2008). The neural correlates of third-party punishment. *Neuron*, *60*, 930–940.

Crick, F. (1994). *The astonishing hypothesis*. New York, NY: Touchstone.

Eshleman, A. (2004). Moral responsibility. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved from plato.stanford.edu/entries/moral-responsibility/

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140.

Gazzaniga, M. S. (2011). *Who's in charge? Free will and the science of the brain*. New York, NY: Harper (Ecco).

Greene, J., & Cohen, J. (2004). For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*, 1775–1778.

Krueger, F., Hoffman, M., Walter, H., & Grafman, J. (in press). An fMRI investigation of the effects of belief in free will on third-party punishment. *Social Cognitive and Affective Neuroscience*.

Monterosso, J., & Schwartz, B. (2012, July 27). Did your brain make you do it? *New York Times*. Retrieved from http://www.nytimes.com/2012/07/29/opinion/sunday/neuroscience-and-moral-responsibility.html?_r=1

Nahmias, E. (2011, November 13). Is neuroscience the death of free will? *New York Times*. Retrieved from http://opinionator.blogs.nytimes.com/2011/11/13/is-neuroscience-the-death-of-free-will/

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

Nelissen, R. M. A., & Zeelenberg, M. (2009). Moral emotions as determinants of third-party punishment: Anger, guilt, and the functions of altruistic sanctions. *Judgment and Decision Making*, *4*, 543–553.

Nichols, S. (2011). Is free will an illusion? *Scientific American Mind*, *22*, 18–19.

Osborne, L. (2003, June 22). Savant for a day. *New York Times*. Retrieved from http://www.nytimes.com/2003/06/22/magazine/22SAVANT.html?pagewanted=all

Paulhus, D. L., & Carey, J. M. (2011). The FAD-Plus: Measuring lay beliefs regarding free will and related constructs. *Journal of Personality Assessment*, *93*, 96–104.

Paulhus, D. L., & Margesson, A. (1994). *Free Will and Determinism (FAD) scale*. Unpublished manuscript, Department of Psychology, University of British Columbia, Vancouver, Canada.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*, 717–731.

Rai, T. S., & Holyoak, K. J. (2013). Exposure to moral relativism compromises moral behavior. *Journal of Experimental Social Psychology*, *49*, 995–1001.

Schooler, J. W. (2010). What science tells us about free will. In R. F. Baumeister, A. R. Mele, & K. D. Vohs (Eds.), *Free will and consciousness: How might they work?* (pp. 191–218). Oxford, England: Oxford University Press.

Shariff, A. F., Schooler, J., & Vohs, K. D. (2008). The hazards of claiming to have solved the hard problem of free will. In J. Baer, J. C. Kaufman, & R. F. Baumeister (Eds.), *Are we free? Psychology and free will* (pp. 181–204). Oxford, England: Oxford University Press.

Smith, A. (1759). *Theory of moral sentiments*. London, England: A. Miller.

Soon, C., Brass, M., Heinze, H., & Haynes, J. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, *11*, 543–545.

Stephen, S. J. F. (1883). *A history of the criminal law of England* (Vol. 3). New York, NY: Macmillan.

Tonry, M. (2004). *Thinking about crime: Sense and sensibility in American penal culture*. New York, NY: Oxford University Press.

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.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validations of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070.

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

Wolfe, T. (1997, February 2). Sorry, but your soul just died. *Forbes*. Retrieved from http://90.146.8.18/en/archiv_files/19971/E1997_236.pdf