

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

Azim F. Shariff*
University of Oregon

Joshua D. Greene
Harvard University

Johan C. Karremans
Behavioral Science Institute, Radboud University Nijmegen

Jamie Luguri
Yale University

Cory Clark
University of California, Irvine

Jonathan W. Schooler
University of California, Santa Barbara

Roy F. Baumeister
Florida State University

Kathleen D. Vohs
University of Minnesota

***Correspondence to:**

Azim F. Shariff
1227 University of Oregon
Eugene, OR 97401
(541) 222 9354
shariff@uoregon.edu

Cite as: Shariff, A.F., Greene, J.D., Karremans, J.C., Luguri, J., Clark, C. J., Schooler, J.W.,
Baumeister, R.F. & Vohs, K.D. (in press). Free Will and Punishment: A
Mechanistic View of Human Nature Reduces Retribution. *Psychological
Science*.

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 using both measured and manipulated free will beliefs found that people with weaker beliefs reported less retributive, but not consequentialist, punishment towards criminals (Study 1). Subsequent studies showed that exposing people to research about the neural bases of human behavior, either through lab-based manipulations or by virtue of having taken an undergraduate neuroscience class, reduced retributive punishment (Studies 2-4). These results illustrate the consequences that exposure to debates about free will and scientific research on the neural basis of behavior may have on attributions of moral responsibility.

Keywords: Free Will, Punishment, Morality, Responsibility, Blame

Most people believe that humans have free will (Nahmias, Morris, Nadelhoffer, & Turner, 2005). However, longstanding controversies remain over what forms of free will can actually exist alongside the known laws of nature (see Baer, Kaufman & Baumeister, 2008). Although few deny that humans regularly make uncoerced choices and exercise self-control, many scientists and philosophers have taken issue with an idea of free will whereby 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 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 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 evermore mechanical causes for human behavior (Greene & Cohen, 2004; Nahmias, 2011; Wolfe, 1997), questions have emerged about the psychological and societal consequences of reduced free will beliefs (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 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 their damage 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 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, research has found that, compared to others, participants whose free will beliefs were experimentally diminished were less helpful and more likely to lie, cheat, steal, and act aggressively (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.

The current work considers 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 incur 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: those who harm others should be harmed themselves. Retributivism regards reciprocity as moral justification for punishment, independent of any other benefits the punishment may bring. Retributivism, in other words, holds that the

95 point of punishment is to extract suffering from the transgressor as ‘just deserts.’
96 Retributivism is contrasted with consequentialist justifications for punishment,
97 which depend upon the utilitarian benefits of punishment (such as rehabilitation or
98 deterrence). Consequentialism selects punishment based on what has the best social
99 consequences, regardless of how much or how little the transgressor suffers (if at
100 all) or indeed whether the transgressor even deserves punishment.

101 Punishment has been correlated with beliefs in free will. For instance,
102 Krueger and colleagues found those with higher free will belief also tended to be
103 more punitive, though only for transgressions that were not especially emotionally
104 arousing (Krueger, Hoffman, Walter & Grafman, in press). Theorists have argued
105 however that only retributivist, and not consequentialist, motivations for
106 punishment should depend on the strong and embattled form of free will discussed
107 above (Greene & Cohen, 2004). Since consequentialism is unconcerned with
108 whether someone deserves punishment in an ultimate sense, and instead focused
109 on whether punishment would have a positive utilitarian effect, it is not threatened
110 by mechanistic arguments against free will. In contrast, because retributivism is
111 based on transgressors being blameworthy and deserving suffering due to their
112 transgressions, anything that diminishes that sense of deservingness—such as an
113 inability to have freely chosen not to commit the transgression—diminishes the
114 justification for retributive punishment. Though these are theoretical connections,
115 there is evidence to suggest that people do base their moral decision-making, in
116 part, on their philosophical worldview (e.g., Rai & Holyoak, 2013; Vohs & Schooler,
117 2008). Therefore, we predict 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

We tested the relationship between free will beliefs and attitudes towards consequentialist and retributivist punishment (Study 1). We then moved to experimental methods to test the effects on punitive attitudes of diverse manipulations aimed at diminishing free will beliefs. In some experiments, we had participants read about challenges to free will whereas others exposed participants to research about the neural mechanisms underlying human action. The outcome measures in all studies were retributivist attitudes towards criminality (Studies 2-4). Throughout all studies, we predicted that reduced belief in free will would predict weaker retributivist punishment.

Study 1: Correlations between Free Will Beliefs and Retribution

Study 1 tests 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 aims to benefit society, we predicted that retributive punishment would positively relate to free will beliefs, whereas consequentialist punishments would be unrelated.

Method

Two hundred and forty-four Americans (147 female, $M_{\text{age}} = 36.81$) participated online via Amazon's Mechanical Turk. Respondents completed the seven-item Free Will subscale of the Free Will Determinism Plus scale (FAD+; Paulhus & Carey, 2010; $\alpha = .88$. Items—e.g. "People have complete control over the decisions they make"—were rated from 1 (*Strongly disagree*) to 5 (*Strongly agree*)). The FAD+ has 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 towards retributivist and consequentialist motivations for punishment, participants first read descriptions of retributivism and consequentialism as motivations for punishment, and then indicated on two separate 1 to 7 likert scales how important retributivism and consequentialism should be in determining motivations for criminal punishment (Supplementary Materials). Participants also completed questions about their education, religiousness, political ideology for economic and social issues, and demographics.

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, the retributivism and consequentialism scales were regressed on Free Will subscale scores. As predicted, stronger belief in free will predicted higher scores on retributive punishment ($\beta =$

.24, $p < .001$), but were not predictive of supporting consequentialist punishment, $\beta = -.02$, $p = .72$. Effects remained significant when statistically controlling for age, gender, education, religiosity, economic and social political ideology (Table 1).

Study 1 supported hypotheses 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 cannot determine whether a reduction of free will beliefs would lead to a resultant reduction in retributivism. As a result, Studies 2-4 moved to experimental methods to causally test the relationship between free will beliefs and retributive punishment by manipulating free will beliefs directly.

Study 2: Manipulated Free Will Beliefs and Retributivism

Method

Participants. Forty-six students (30 female, $M_{\text{age}} = 20.44$) participated for partial course credit.

Procedure. Under the guise of unrelated experiments, participants first completed a free will manipulation. Participants were randomly assigned to read 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 (Supplementary Materials). In the neutral condition, the passage was unrelated to free will. This task has been previously validated by Vohs & Schooler (2008), and testing the manipulation in an independent sample revealed that the

passages led to differences in free will beliefs, as measured by a single item (Supplementary Materials), in the expected direction, $t(205) = 2.55, p = .01$.

Next, participants read a fictional vignette involving an offender who beat a man to death (Supplementary Materials). Acting as hypothetical jurors, participants recommended the length (if any) of a prison sentence that would follow a nearly 100% effective two-year rehabilitation treatment. The notion that the offender had been rehabilitated was used in order to isolate participants' desire for punishment as retribution¹. To further focus participants on retributive, rather than consequentialist punishment, participants read that the prosecution and defense had agreed that the rehabilitation would prevent recidivism and, further, that extant evidence indicated that detention after rehabilitation does not work to deter other potential criminals.

Participants then chose among seven punishment options: 1 = Treatment time only, no imprisonment; 2 = 2 years imprisonment post-treatment (all subsequent years indicate post-treatment imprisonment years), 3 = 5 years; 4 = 10 years; 5 = 25 years with the chance of parole after 15; 6 = 25 years with no chance of parole; 7 = Life imprisonment with no chance of parole. Last, participants completed a demographics questionnaire and suspicion probe (Supplementary Materials).

Results

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

As predicted, participants exposed to the anti-free will passage recommended significantly lighter prison sentences than participants exposed to the neutral passage, $M = 2.91$, $SD = 1.08$ vs. $M = 3.96$, $SD = 1.49$; $t(44) = 2.71$, $p < .05$, Cohen's $d = .82$ (Figure 1). Translating these means to their corresponding sentence lengths, participants whose free will beliefs had been experimentally diminished recommended roughly half the length of imprisonment (~5 years) relative to neutral participants (~10 years).

Therefore, Study 2 demonstrated that experimentally diminishing free will beliefs alters legal judgments by reducing inclinations for retributive punishment. Study 3 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 moral responsibility judgments.

Study 3: Systematic Exposure to Neuroscience and Retributivism

Study 2's anti-free will manipulation came from a quite opinionated essay written by a Nobel laureate (Crick, 1994). Study 3 used a subtler manipulation that may represent how the free will beliefs of everyday people 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 that implied 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 that implies a mechanistic view of human action would reduce belief in free will and likewise retributive tendencies. Moreover, we predicted that exposure to mechanistic neuroscience would reduce the extent to which a transgressor is perceived as blameworthy, and these assessments would mediate the effect of condition on sentencing recommendations.

Method

Participants. Ninety-one undergraduates participated for partial course credit. Four participants were excluded—three for suspicion, one for admitted intoxication—leaving 88 participants (61 female, $M_{\text{age}} = 20.81$).

Procedure. Under the guise of unrelated experiments, participants first read two popular science articles made to look as if they were from *Scientific American* and *New Scientist* websites (Supplementary Materials). 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). Neutral condition participants 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, pre-testing with an independent sample revealed that the neuroscience passages led to lower free will beliefs than did the neutral science passages, $t(198) = 2.35, p = .020$. This confirms that relative to other scientific descriptions, exposure

to discussions of scientific research promoting a mechanistic view of human behavior diminishes belief in free will.

For the ostensible second study, participants read the vignette from Study 2 and recommended a prison sentence. In addition, participants rated the offender's blameworthiness (1 = "Not at all deserving of blame"; 7 = "Completely deserving of blame"). Participants then completed the FAD+ as a manipulation check and the Positive and Negative Affect Scale (PANAS; Watson, Clark, & Tellegen, 1988) to account for mood effects that might emerge from challenges to intuitions about the mechanisms behind human action. Last, participants completed a demographics questionnaire and suspicion probe (Supplementary Materials).

Results

Mood. Neither positive nor negative mood significantly differed by condition (p s > 0.4). Higher levels of positive affect corresponded to lower attributions of blame, $r(86) = -.34$, $p = .002$. No other relationships were found. Controlling for mood did not markedly change the main results².

Manipulation check. Consistent with results from the independent manipulation test, participants who read neuroscience articles reported marginally lower belief in free will compared to participants who read neutral articles, $F(1,87) = 3.46$, $p = .07$.

² Allaying concerns that effects were due to mood, the effects of condition on the FAD, punishment and blameworthiness controlling for the PANAS were $F(1,87) = 5.98$, $p = .02$; $F(1,84) = 5.06$, $p = .03$ and $F(1,87) = 5.77$, $p = .04$, respectively.

Punishment and blameworthiness. Consistent with predictions, participants who read neuroscience articles recommended significantly shorter prison sentences than participants that read other science articles ($M = 3.10$, $SD = 1.48$ vs. $M = 3.83$, $SD = 1.77$; $t(86) = 2.09$, $p = .04$, $d = .45$; Figure 1) and blamed the transgressor less ($M = 5.48$, $SD = 1.17$ vs. $M = 6.03$, $SD = 0.92$; $t(86) = 2.40$, $p = .02$, $d = .52$). Punishment and blameworthiness were significantly correlated, $r(88) = .37$, $p < .001$.

Mediation. We predicted that reading neuroscience articles would decrease participants' perception that the offender is morally blameworthy, which would account for recommendations of lighter sentences. Consistent with predictions, bootstrapping analyses (10,000 resamples; Preacher & Hayes, 2004) indicated that blameworthiness mediated the effect of condition on sentencing recommendations (95% CI = -0.68 to -0.03; Figure 2).

Study 3 bolstered Study 2's conclusions in several ways. First, it replicated Study 2's finding that experimental manipulations that decrease free will beliefs also reduce retributive punishment. Second, the mediational effect on 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 press at *Psychological Science*

Study 4 delved further into this issue of naturalistic manipulations of free will beliefs. Instead of using laboratory-based manipulations, it employed a naturalistic method by which people learn about mechanistic causes underlying human action, namely university neuroscience classes.

In Study 4, 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, ten weeks apart. We predicted that learning about the brain would reduce retributive punishments, conceptually replicating 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 non-neuroscience class as a neutral comparison.

Method

Participants. Students from an introductory cognitive-neuroscience class, ($n = 34$) and an introductory geography class ($n = 36$) participated in exchange for a raffle prize. Two students who were simultaneously taking both classes were dropped from analysis, leaving 68 participants (39 female, $M_{\text{age}} = 20.44$).

Procedure. On the first day of class, students completed a shortened version of the punishment scenario from Studies 2-3 (Supplementary Materials). Students were 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 anticipated class standing. Instructors for both classes were blind to the hypothesis.

Results

Supporting predictions, a paired-sample comparison showed that prison sentence recommendations decreased from the start to end of the neuroscience class, from 3.41 to 2.91 on a 7-point scale, $t(33) = 2.15$, $p = .04$, $d = .44$. No change was found for students in the geography class, $M_{T1} = 3.32$ vs. $M_{T2} = 3.08$, $t(33) = .94$, $p = .41$, though the spreading interaction between the classes did not reach statistical significance, $F(1,68) = .22$, $p = .64$.

Were changes in retributivism related to student interest and learning outcomes? In neither class did grades predict changes in sentence recommendations. However, for students in the neuroscience course, the degree to which sentence recommendations decreased was strongly correlated with an increase from students' self-reported knowledge of the brain during the first class to what they reported at 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 ($r = .02$, $p = .90$). These relationships significantly differed between classes (Steiger's $Z = 2.00$, $p = .05$), suggesting 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 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 that implied a mechanistic basis to human action—either through reading popular science articles or through taking an introductory college class—similarly produces a reduction in retributivism. These results suggest that shifts in people’s philosophical worldview about free will beliefs, even through just 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 famously wrote that, “The sentence of the law is to the moral sentiment of the public what a seal is to hot wax” (Stephen, 1883, p.423). In recent years, justice researchers and advocates have argued for a switch from retributive to *restorative* justice—a consequentialist response 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 presented in the context of a mechanistic worldview is sufficient to produce these effects (Studies 3-4) 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 has shown that judges afforded shorter sentences to hypothetical psychopathic criminals when criminals' psychopathy included a biomechanical description, compared to when such a description was absent (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 was 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 the participant for being more punitive. While not negating the idea that lowered free will beliefs provide an excuse for self-interested behavior, our data 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, it is notable that 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. Thus, a societal shift away from endorsing free will could occur without disrupting the functional role of punishment, even while eschewing the need for blame and desire for retribution. Society's practical need for law and order could be fulfilled, leaving the social benefits of punishment intact, while avoiding the unnecessary human suffering and the 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—call for understanding.

References

- 390
 391 Aspinwall, L. G., Brown, T. R., & Tabery, J. (2012). The double-edged sword: does
 392 biomechanism increase or decrease judges' sentencing of psychopaths?
 393 *Science*, 337, 846-849.
- 394 Baer, J., Kaufman, J. C., & Baumeister, R. F. (Eds.). (2008). Are we free? Psychology
 395 and free will. Oxford University Press, USA.
- 396 Bargh, J.A. (2008). Free will is unnatural. In J. Baer, J. Kaufman, & R. Baumeister
 397 (Eds.), *Are we free? The psychology of free will* (pp. 128-154). New York:
 398 Oxford University Press.
- 399 Baumeister, R. F., Masicampo, E. J., & DeWall, C. N. (2009). Prosocial benefits of
 400 feeling free: Disbelief in free will increases aggression and reduces
 401 helpfulness. *Personality and Social Psychology Bulletin*, 35(2), 260-268.
- 402 Braithwaite, J. (2002). *Restorative Justice & Responsive Regulation*. Oxford University
 403 Press.
- 404 Buckholtz, J.W., Asplund, C.L., Dux, P.E., Zald, D.H., Gore J.C., Jones, O.D., & Marois, R.
 405 (2008). The neural correlates of third-party punishment. *Neuron*, 60, 930-40.
- 406 Crick, F. (1994). *The astonishing hypothesis*. New York: Touchstone.
- 407 Eshleman, A. (2004). Moral responsibility. In E. N. Zalta (Ed.), *The Stanford*
 408 *encyclopedia of philosophy*. Palo Alto, CA: Stanford University.
- 409 Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137.

- 410 Gazzaniga, M.S. (2011). *Who's in charge? Free will and the science of the brain*. New
411 York: Harper (Ecco).
- 412 Greene, J., & Cohen, J. (2004). For the law, neuroscience changes nothing and
413 everything. *Philosophical Transactions of the Royal Society of London B*, 359,
414 1775-1778.
- 415 Krueger, F., Hoffman, M., Walter, H., & Grafman, J. (in press). An fMRI investigation of
416 the effects of belief in free will on third-party punishment. *Social cognitive
417 and affective neuroscience*.
- 418 Monterosso, J., & Schwartz, B. (2012, July 27). Did your brain make you do it. *New
419 York Times*. Retrieved from:
420 [http://www.nytimes.com/2012/07/29/opinion/sunday/neuroscience-and-
421 moral-responsibility.html?_r = 1](http://www.nytimes.com/2012/07/29/opinion/sunday/neuroscience-and-moral-responsibility.html?_r=1)
- 422 Nahmias, E. (2011, November 13). Is neuroscience the death of free will? *New York
423 Times*. Retrieved from:
424 [http://opinionator.blogs.nytimes.com/2011/11/13/is-neuroscience-the-
425 death-of-free-will/](http://opinionator.blogs.nytimes.com/2011/11/13/is-neuroscience-the-death-of-free-will/)
- 426 Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2005). Surveying freedom: Folk
427 intuitions about free will and moral responsibility. *Philosophical Psychology*,
428 18, 561-584.

- 429 Nelissen R. M. A., Zeelenberg M. (2009). Moral emotions as determinants of third-
430 party punishment: anger, guilt, and the functions of altruistic sanctions.
431 *Judgment & Decision Making*, 4, 543–553.
- 432 Nichols, S. (2011). Is free will an illusion? *Scientific American Mind*, 22, 18-19.
- 433 Osborne, L. (2003, June 22). Savant for a day. *New York Times*. Retrieved from
434 <http://www.nytimes.com/2003/06/22/magazine/22SAVANT.html?pagewanted=all>.
435 nted = all.
- 436 Paulhus, D. L., & Carey, J. M. (2011). The FAD-Plus: Measuring lay beliefs regarding
437 free will and related constructs. *Journal of Personality Assessment*, 93, 96-104.
- 438 Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating
439 indirect effects in simple mediation models. *Behavior Research Methods*,
440 *Instruments, & Computers*, 36, 717-731.
- 441 Rai, T. S., & Holyoak, K. J. (2013). Exposure to Moral Relativism Compromises Moral
442 Behavior. *Journal of Experimental Social Psychology*, 49, 995-1001.
- 443 Schooler, J.W. (2010) What Science Tells Us About Free Will. In R. F. Baumeister, A.
444 R. Mele & K. D. Vohs (Eds.) *Free Will and Consciousness: How Might They*
445 *Work?* (pp. 191-218). Oxford University Press.
- 446 Shariff, A. F., Schooler, J., & Vohs, K. D. (2008). The hazards of claiming to have
447 solved the hard problem of free will. In J. Baer, J. C. Kaufman, & R. F.
448 Baumeister (Eds.), *Are we free? Psychology and free will* (pp. 181–204). New
449 York : Oxford University Press.

- 450 Smith, A. (1759). *Theory of moral sentiments*. London: A. Miller.
- 451 Soon, C., Brass, M., Heinze, H., & Haynes, J. (2008). Unconscious determinants of free
452 decisions in the human brain. *Nature Neuroscience*, 11, 543-545.
- 453 Stephen, S. J. F. (1883). *A history of the criminal law of England* (Vol. 3). Macmillan.
- 454 Tonry, M. (2004). *Thinking about crime: sense and sensibility in American penal*
455 *culture*. New York: Oxford University Press.
- 456 Vohs, K.D., & Schooler, J. W. (2008). The value of believing in free will: Encouraging a
457 belief in determinism increases cheating. *Psychological Science*, 19, 49-54.
- 458 Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validations of brief
459 measures of positive and negative affect; The PANAS scales. *Journal of*
460 *Personality and Social Psychology*, 54, 1063-1070.
- 461 Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.
- 462 Wolfe, T. (1997, February 2). Sorry, but your soul just died. *Forbes*. Retrieved from:
463 http://90.146.8.18/en/archiv_files/19971/E1997_236.pdf.
- 464 AS, JG, JK, JS & KV developed the study concept. AS & JL conducted the studies. AS
465 analyzed and interpreted the data. AS drafted the paper, and CC, JG, JK, JL , JS, KV and
466 RB provided critical revisions. All authors approved the final version of the paper
467 for submission.

468

469

Authors' Note

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

474

Table

Table 1: Retributive punishment, consequentialist punishment, and desire for vengeance regressed on free will belief, age, sex, educational attainment, religiosity, and political ideology.

	Retributive Punishment	Consequentialist Punishment
Free Will Belief (no controls)	.242***	-.024
Free Will Belief (with controls)	.164*	.018
Age	.088	.156*
Sex (Male)	.050	.045
Education	-.139*	.056
Religiosity	-.158*	.034
Political Ideology (Social)	.323***	-.215*
Political Ideology (Economic)	.003*	-.013

Values indicate standardized beta coefficients.

Higher values for political ideology indicate more conservative.

* $p < .05$, ** $p < .01$, *** $p < .001$

Figure Captions:

Figure 1. Mean punishment recommendations across conditions. Exposure to anti-free will arguments (Study 2) and neuroscientific research (Study 3), compared to neutral passages, reduced the recommended prison sentence for a hypothetical criminal.

Figure 2. Exposure to neuroscientific research reduced punishment recommendations (Study 3). This effect is mediated by the perceived blameworthiness of the offender. *a*, *b* and *c* denote the standardized regression coefficients. *c'* represents the mediated effect of condition on punishment, taking into account the indirect effect of blameworthiness. * denotes $p < .05$; ** denotes $p < .01$.