Consciousness and Cognition 75 (2019) 102808

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Behavioral explanations reduce retributive punishment but not reward: The mediating role of conscious will

Joshua A. Confer*, William J. Chopik

University of California, Berkeley, United States
Michigan State University, United States

ARTICLE INFO

Keywords:
Free will
Responsibility
Punishment
Retribution
Reward
Morality

ABSTRACT

Conceptions of responsibility are associated with the degree to which people ascribe conscious will to others. However, it is not known how the biological and environmental circumstances of moral actors independently impact attributions of conscious will. Although reductions in conscious will are associated with diminished punishment of criminals, does a reduced sense of conscious will of a hero affect support for reward? In two pre-registered studies (total $N = 2668$), we investigated the effects of biological or environmental histories on judgements of punishment and reward. Biological and environmental circumstances (especially biological) reduced perceptions of conscious will, which in turn reduced conviction judgements and punishment severity (Studies 1–2). In the context of a moderated mediation, we found that reductions in perceptions of conscious will were unrelated for a desire to reward a hero (Study 2). Findings are discussed in the context of a model of judgement, conscious will, and responsibility.

1. Introduction

The same act does not prompt uniform judgements of moral responsibility across individuals. Evaluations of moral responsibility are partially constructed based on assessments of how much conscious control a moral actor has over an event. However, the extent to which specific knowledge of an individual's biological makeup and environmental history impact perceiver's attributions of will and responsibility are not well understood. In the current studies, we manipulated the biological and environmental histories of a criminal and a hero to examine how these considerations affect judgements of conscious will, punishment, and reward.

In the following sections, we provide an overview of conscious will, responsibility, and moral judgements. Specifically, we focus on the functions that these beliefs and judgements serve, how they might be malleable to biological and environmental circumstances, and how judgements of transgressions and virtuous acts might differ from one another.

1.1. Conscious will

The focus of the current report is not to speculate into the philosophical truth of free will, or the conscious ability to do otherwise, but rather examine the consequences of people's folk beliefs about free will for judgement and decision making (Nahmias, Morris, Nadelhoffer, & Turner, 2005).

At a basic level, people often seek to uncover reasons *why* someone behaves in a particular manner (Ross, 1977). This process usually involves inferences about the target's mental states, such as their beliefs, desires, and intentions (Fodor, 1987; Premack &

* Corresponding author at: Department of Psychology, University of California-Berkeley, 2121 Berkeley Way, Berkeley, CA 94704, United States.
E-mail address: confer@berkeley.edu (J.A. Confer).

https://doi.org/10.1016/j.concog.2019.102808
Received 30 August 2018; Received in revised form 30 July 2019; Accepted 22 August 2019

Woodruff, 1978). People even project mental states onto non-sentient objects, inferring that geometric shapes move with volition in the same way people do (Heberlein & Adolphs, 2004). However, when considering the mental states and behavior of other humans, people do not simply consider whether a person possesses a particular mental state or committed an action. Rather, people also distinguish between mental states and actions that are freely and consciously willed (e.g., cheating on a test) or constrained and not consciously willed (e.g., seizures).

In general, most lay people believe that individuals possess conscious control over their mental states and actions (Cusimano & Goodwin, 2019; Monroe & Malle, 2010; Nichols, 2004). Moreover, Sarkissian et al. (2010) has found belief in free will across several cultures, which might suggest that free will is a near universal tenet. Yet, when evaluating human mental states and behavior, neither observers nor targets have complete insight into the processes that govern them (Nisbett & Wilson, 1977; Spinoza, 1985). Without this insight, it follows that the exact amount and influence of conscious will on observable actions is unknowable. In addition, the total effect of one's biological makeup and environmental history on this will is also unknowable. For example, there is no way of measuring the exact extent that an environmental circumstance (e.g. poverty) affects an individual's mental state or action in a given context—let alone how an amalgamation of multiple biological and environmental circumstances impacts mental states or behavior. When biological and environmental circumstances are viewed to cause a behavior, the lay assumption might be that at least some of the causal chain occurs outside of the conscious self's decision. Indeed, it has been proposed that free will is often partially inferred by observing constraints (e.g., a person was "forced to do something"), or the lack thereof, on others' behavior (Nichols, 2004). Nevertheless, people make assessments about mental states and behavior on a spectrum ranging from internally to externally caused. Such a spectrum has large implications for how we morally evaluate others. Although the exact degree to which biological and environmental circumstances compromise conscious will is unknowable, in the current studies we specifically test whether such circumstances reduce *perceptions* of conscious will among laypeople.

1.2. *Transgressions*

1.2.1. *Holding others morally responsible*

The ubiquity of free will beliefs might partially explain the mechanisms through which we hold others responsible for their behavior—primarily through how we judge and punish moral transgressions.

Indeed, the concepts of free will and moral responsibility are tightly intertwined. Attributions of responsibility appear to be altered if a target is viewed to not have conscious control over their actions (Shariff et al., 2014). Whether an individual possesses free will likely does not affect whether they are held *causally* responsible (actions still lead to effects), but it may affect perceptions of whether the individual is held *morally* responsible. As Shariff et al. (2014) explain, insinuating that an individual *should* have done otherwise implies one *could* have done otherwise.

People clearly recognize that the behavior of others is not always spontaneously willed. Instead, people likely assume that individuals can be influenced by various biological or environmental circumstances. After this attribution, there is then an unknown proportion of one's will "left over" which is not viewed to be influenced or constrained by these circumstances. The residual conscious will may then be used to form direct judgements about moral responsibility.

Previous research demonstrates a consistent connection between free will and responsibility in the context of a negative act. Collectively, numerous studies suggest that perceptions of a target's moral responsibility are mitigated when a behavior is revealed to be derivative of either particular biological circumstances (e.g. neurological information, genetics, reflexes) or environmental circumstances (e.g. abuse, poverty, crime of passion) (Alicke, 2000; Aspinwall, Brown, & Tabery, 2012; Gray, Young, & Waytz, 2012; Provencher & Fincham, 2000; Steinberg & Scott, 2003). However, the degree to which reductions in responsibility from these circumstances are explained by attributions of conscious will (and their effects on punishment decisions) has not been explicitly measured in a formal process model.

1.2.2. *The influence of biological versus environmental circumstances*

Discussions of biological and environmental circumstances are often found in court cases and dramatized in popular media. Defense attorneys are increasingly including biological information in a variety of criminal cases, like capital murder, by using neuroscientific and genetic findings (Bernet, Vnencak-Jones, Farahany, & Montgomery, 2007; Farahany & Coleman, 2006). However, when considering influences on one's conscious will, it has been documented that criminologists are partial to environmental circumstances and generally discount the role of biology (Cooper, Walsh, & Ellis, 2010; Walsh, 2010). How are lay people's views of conscious will differentially influenced by these circumstances? In other words, are they more swayed by biological or environmental circumstances when making judgements about moral responsibility?

Although previous research is limited, findings may suggest that biological circumstances play a greater role in evaluations of behavior compared to environmental circumstances. Dar-Nimrod and Heine (2011) document that people tend to explain behavior in accordance with learned genetic information (biological), ignoring other factors that influence behavior (e.g., social upbringing, environmental). This follows from the fact that genes are viewed as proximate and determined causes of behavior (Meehl, 1977). Environmental factors may require an additional level of abstraction (e.g., environmental circumstances affect biological circumstances which then affect behavior), and therefore have a smaller, more distal influence relative to biological factors. Indeed, it has been consistently demonstrated that social factors are downplayed in favor of inherent psychological dispositions (Gilbert & Malone, 1995). Therefore, it may be that individual dispositions that are perceived to follow from biological circumstances (e.g., genetics or neural development) are judged to produce greater constraints on an individual's will compared to social, environmental circumstances (e.g., abuse, upbringing). However, it is unclear how biological and environmental circumstances influence perceptions of conscious will. Variation in these influences may be important, as a bias in these perceptions may lead to errors in judgements of how a behavior occurred and affect subsequent interpretations of that behavior. The current study hopes to address the question of whether biological or environmental circumstances pose greater constraints on perceived conscious will.

In summary, diminished perceptions of an individual's will may alleviate their moral responsibility—trespassers seem to transition from agents with complete control to victims of circumstance (biological or environmental). However, rarely do researchers explicitly test reductions in the ability to do otherwise, or conscious will, as a *mediator* of the link between biological and environmental circumstances on responsibility and punishment decisions. That is, do biological and environmental circumstances reduce perceptions of conscious will and thus reduce the likelihood of punishing? Further, the effects of biological and environmental circumstances on responsibility and punishment are always tested in isolation of one another. Thus, the extent and relative influence to which biological versus environmental circumstances affect lay evaluations of conscious will, and in turn responsibility and punishment, remains unclear. Studies 1 and 2 tested these questions directly.

1.3. Reward

Moral wrongdoers are not the only group of people to whom we ascribe responsibility. It would seem that if transgressors are responsible for their immoral actions, those who commit virtuous actions would also be responsible for them. Less is known about the association between views of conscious will and responsibility in the context of positive acts. Previous research has indicated that moral trespassers garner larger attributions of responsibility compared to virtuous actors (i.e., we judge moral transgressors as more responsible for their negative actions and moral paragons as less responsible for their positive acts). Indeed, compared to positive acts, negative actions draw more attention and demand more of an explanation towards the cause of the act (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Taylor, 1991). In line with this, individuals who commit negative acts receive higher judgements of free will and responsibility compared to individuals who commit positive acts (Feldman, Wong, & Baumeister, 2016; Hamlin & Baron, 2014; Knobe, 2003; Morewedge, 2009). Worth noting, individuals who commit positive actions *are* given higher responsibility and conscious will attributions compared to those who commit morally neutral actions (Clark et al., 2014; Clark, Shniderman, Luguri, Baumeister, & Ditto, 2018). This implies that heroes are given some portion of responsibility. Together, this research suggests that moral paragons are viewed as possessing free will and responsibility, albeit less so than moral transgressors.

Are a hero's actions viewed differently if the behavior arises from a source other than the actor's own conscious will (as may be the case with criminals)? Clark et al. (2018) suggests that there may be a positive relationship between higher conscious will attributions and reward judgements, such that we give higher rewards to heroes who have conscious will. However, this possibility has yet to be experimentally investigated. Given that diminished judgements of conscious will reduce a desire to punish criminals (Shariff et al., 2014), does it also predict a decrease in desire to reward heroes? As previously discussed, punishment involves a higher allocation of responsibility compared to neutral and positive acts. Thus, it could be that reductions of responsibility have less bearing on reward decisions compared to punishment decisions. Study 2 investigated the effect of manipulating perceptions of conscious will to examine the consequences on reward judgements. Further, Study 2 also examined whether biological and environmental constraints on conscious will differentially affect judgements of reward.

1.4. The current studies

There are many unknowns about the roles of biological or environmental circumstances for reducing conscious will and the consequences of these reductions for reward and punishment decisions. In the current studies, we address many of these gaps. Specifically, we compare the relative influence of biological or environmental circumstances for reducing perceptions of conscious will. We formally model the process through which biological or environmental circumstances affect reward and punishment decisions (e.g., via reductions in conscious will). Finally, we examine conscious will's possible similarities or asymmetries in punishment versus reward decisions.

The scope of an agent's conscious control over an event may have consequences for how they are morally evaluated. If a negative action can be partially explained by either the biological or environmental circumstances of a transgressor, we predict judgements of will and retributive punishment should decline (Studies 1 and 2).

Theoretically, moral actors who commit virtuous actions are also responsible for their behavior. Study 2 investigated whether reductions in perceptions of a hero's conscious will affect subsequent reward decisions. Across both studies, the extent and relative influence of individual biological and environmental circumstances on will and punishment/reward judgements were assessed.

2. Study 1

In Study 1, we examined whether knowledge of biological and environmental circumstances imposed constraints on perceptions of will, and ultimately the relative effects these constraints had on judgements of punishment. Participants imagined that they were serving as jurors on a murder case. We presented participants with vignettes detailing a crime. In each condition, we added explanations of the behavior by providing information about the perpetrator's biological makeup or environmental history. Following Shariff et al.'s (2014) method, all conditions explained that social benefits (i.e. rehabilitation and deterrence) would be implemented in order to isolate judgements of retribution from consequentialist considerations.[1] We hypothesized that biological and environmental circumstances would reduce lay perceptions of conscious will which would in turn reduce the desire to elicit retribution.

2.1. Method

2.1.1. Participants

One thousand, one hundred and one ($N = 1101$) respondents ($M_{age} = 35.87$, $SD = 10.79$, 41.9% Female) were recruited from Amazon Mechanical Turk (MTurk). Participants received $1.00 for participating in the study. Racial/ethnic composition of the study was as follows: 72.3% white/Caucasian, 14.2% black/African American, 6.3% Asian, and 7.2% other ethnicities. An additional fifty-eight participants were excluded for saying they did not take the study seriously. See the OSF site for pre-registration details regarding the analytic plan and analysis (https://osf.io/ctpb4/). An a priori analysis based on effect sizes from pilot studies (see supplementary materials; $d = 0.55$) suggested a recommended sample size of 176 (with 80% power at $\alpha = 0.05$). However, we wanted to ensure that we could reliably detect an effect *and* estimate its effect size with more precision. Thus, we chose a smaller effect size ($d = 0.20$), so we aimed to collect at least 969 participants.

2.1.2. Procedure and measures

2.1.2.1. Scenario and punishment. Participants began by reading a short description of a fictional man, William, who was on trial for undisputedly killing a coworker in a shooting. Participants were then assigned to one of three conditions which provided apparent explanations of William's behavior. Two conditions varied with respect to whether there were biological or environmental factors that may have affected William's conscious will (see OSF site for full design details and .qsf files). The two conditions explained that William experienced either: (a) a brain tumor (i.e., biological condition) or (b) an abusive childhood (i.e., environmental condition). These two conditions noted that, according to a hospital examination, the circumstance affected William's ability to regulate his motivations and emotions which may increase the likelihood of outbursts of aggression. A third condition acted as a control and did not provide any additional information beyond the basic facts of the case. For summary, the three conditions were as follows:

(1) William killing a coworker; possessing a brain tumor in areas that, according to the hospital physician, affect people's ability to regulate motivations and emotions which can lead to outbursts of aggression; undergoing a 100% effective treatment to prevent future effects of the brain tumor; and being on trial.[2]

(2) William killing a coworker; having a history parental abuse, which according to a hospital physician, affects people's ability to regulate their motivations and emotions which can to outbursts of aggression; undergoing a 100% effective treatment to prevent future effects of the abuse; and being on trial.

(3) William killing a coworker, with no explanations of behavior, undergoing a 100% effective treatment for his minor injuries, and being on trial.[3]

[1] For a brief summary of this distinction, retributive punishment focuses on *deservingness*, and is justified if an individual commits a crime without significant influence from outside circumstances (e.g., they committed a crime and had complete control over their actions). Therefore, judgements of deserved retribution should be influenced by attitudes towards free will (e.g., if they had free will, they should be punished). Indeed, belief in free will is associated with support for retributive punishment (Caspar et al., 2017; Kraus et al., 2013; Shariff et al., 2014). In contrast, consequentialist punishment is less concerned with whether punishment is deserved, but rather the societal benefits that will come from imprisonment, such as security, psychiatric help, and deterrence of future crimes (e.g., if the transgressor is unable harm others in the future and their punishment would not deter others, they should not be punished; Bentham, 1986). Indeed, in a previous iteration of this paper, we ran a study in which we directly manipulated these factors and found that rehabilitation and deterrence reduced punishment decisions ($d = 0.46$; see Supplementary Materials).

[2] We asked participants the extent to which they believed the treatment facility would be effective (for those conditions that read about the rehabilitation). Participants responded on a sliding scale ranging from 0 to 100. The means for Studies 1 ($M = 66.49$, $SD = 26.86$; one-sample t-test (v. 50%): $t(1096) = 20.33, p < .001$) and 2 ($M = 72.45$, $SD = 25.51$; one-sample t-test (v. 50%): $t(1555) = 34.72, p < .001$) were both above the midpoint, suggesting that most participants felt that it was a believable possibility. Worth noting, we ran additional analyses controlling for this variable (for those conditions that had valid data) and the results were virtually the same. Thus, participants mostly found the manipulation to be believable (and it did not account for our findings), but we still acknowledge that such a program and facility may still strike people as far-fetched.

[3] The decision for the treatment of William in the absence of any underlying biological or environmental condition was discussed thoroughly between the authors. Undergoing a 100% effective rehabilitation treatment in the control condition to resolve no specific deficit might seem strange to readers. Further, this control treatment did not mention it would reduce the likelihood of re-offense as the rehabilitation would in the two experimental conditions. Such a difference (that the treatment in the experimental conditions would prevent re-offense and that the treatment in the control condition was ambiguous about this point) might yield an uninterpretable effect size given that the conditions are not perfectly balanced. Worth noting, in the first of our non-pre-registered pilot studies, we included this 100% effective rehabilitation treatment (which prevented the likelihood of re-offense) for all three conditions, including the control condition (see supplement). We found that, aligning with the findings of Study 1 here, relative to the control condition (83.6% guilty), participants in the biological ($\text{Exp}(b)_{\text{Conviction}} = 0.40$; $d_{Sentence} = 0.78$. $d_{Agency} = 1.34$) and environmental conditions ($\text{Exp}(b)_{\text{Conviction}} = 0.60$; $d_{Sentence} = 0.59$; $d_{Agency} = 0.51$) convicted William at a lower rate, sentenced him to fewer years in prison, and attributed less agency to him, respectively. Thus, the results comparing the experimental conditions to the control condition were largely consistent whether it is noted that William received a 100% effective treatment (Studies 1 and 2) or received a 100% effective rehabilitation treatment which would prevent the likelihood of re-offense (pilot Study 1).

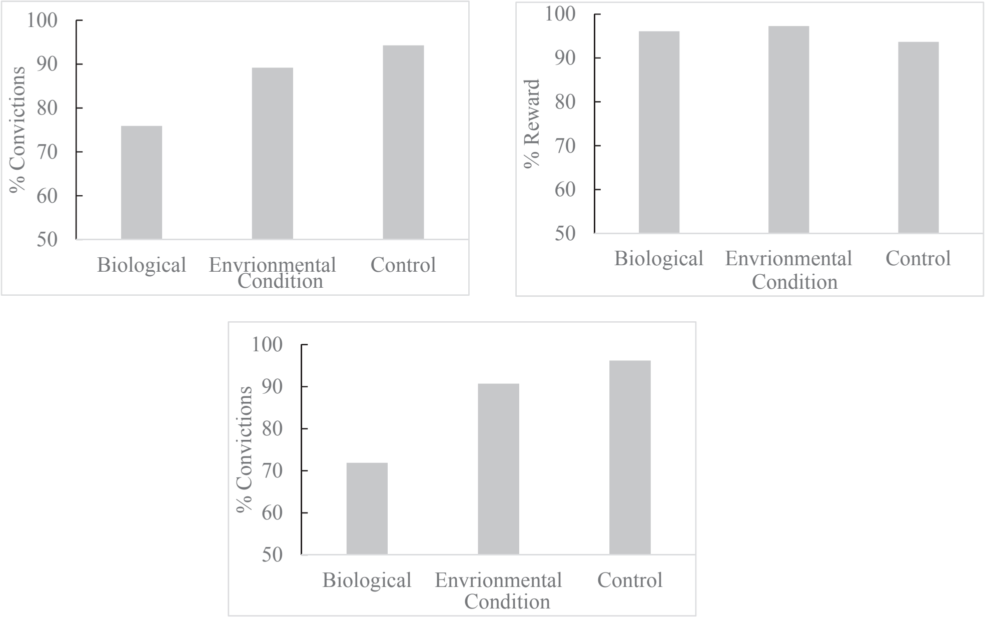


Fig. 1. Decisions by explanation type for punishment (Studies 1 and 2) and reward (Study 2).

All participants were then instructed to decide whether William should serve a prison sentence 1 (*yes*) or 0 (*no*), and if so, the length of this sentence (on a 0 (*no sentence*) to 25 (*25-year sentence*) point scale).

2.1.2.2. Conscious will. Lastly, participants answered a series of questions relating to William's conscious will. The research team created items to reflect various ways in which William’s actions resulted from his own deliberate conscious will (e.g. “Did William have complete conscious control over his actions?“ see OSF site for full measure). Participants responded to each question on a 5-point Likert scale from 1 (*definitely yes*) to 5 (*definitely no*). Responses were reverse coded and averaged such that higher scores reflected greater conscious will ($\alpha = 0.91$).

3. Results

3.1. Biological and environmental influences on conviction and conscious will

3.1.1. Conviction

We ran a logistic regression with biological and environment conditions as dummy codes; the control condition served as reference category. As hypothesized, both the biological ($b = -1.66$, $SE = 0.26$, Wald test = 41.84, $p < .001$, Exp(b) = 0.19) and environmental conditions ($b = -0.70$, $SE = 0.28$, Wald test = 6.15, $p = .01$, Exp(b) = 0.50) reduced conviction rates. Follow-up analyses revealed that convictions rates differed between the biological and environmental conditions as well. Specifically, the conviction rates followed a stepwise pattern, such that convictions were lowest in the biological condition (75.9%), followed by the environmental condition (89.2%), followed by the control condition (94.3%); see the first panel of Fig. 1. Chi-squared analyses confirmed these patterns between the conditions, $\chi^2(2) = 56.31$, $p < .001$.

We conducted one-way analyses of variance (ANOVAs) predicting prison sentence length and conscious will from our three conditions.

3.1.2. Prison sentence

For prison sentence length, there were significant differences between conditions, $F(2, 1098) = 52.71$, $p < .001$. As seen in the left panel of Fig. 2, compared to participants in the control condition, participants in the biological and environmental conditions recommended more lenient prison sentences (all $ps < 0.001$). The three conditions all differed from each other (all $ps < 0.001$) and followed a similar stepwise pattern as seen for convictions: participants in the biological condition gave the most lenient sentences, followed by the environmental condition, followed by the control condition. Thus, reading about either explanation of behavior led to lower conviction rates and more lenient prison sentences (i.e., less punishment).

3.1.3. Conscious will

For conscious will, there were significant differences between conditions, $F(2, 1098) = 98.89$, $p < .001$. As seen in the left panel of Fig. 3, compared to participants in the control condition, participants in the biological and environmental conditions perceived William as having less conscious will (all $ps < 0.001$). The three conditions all differed from each other (all $ps < 0.001$) and

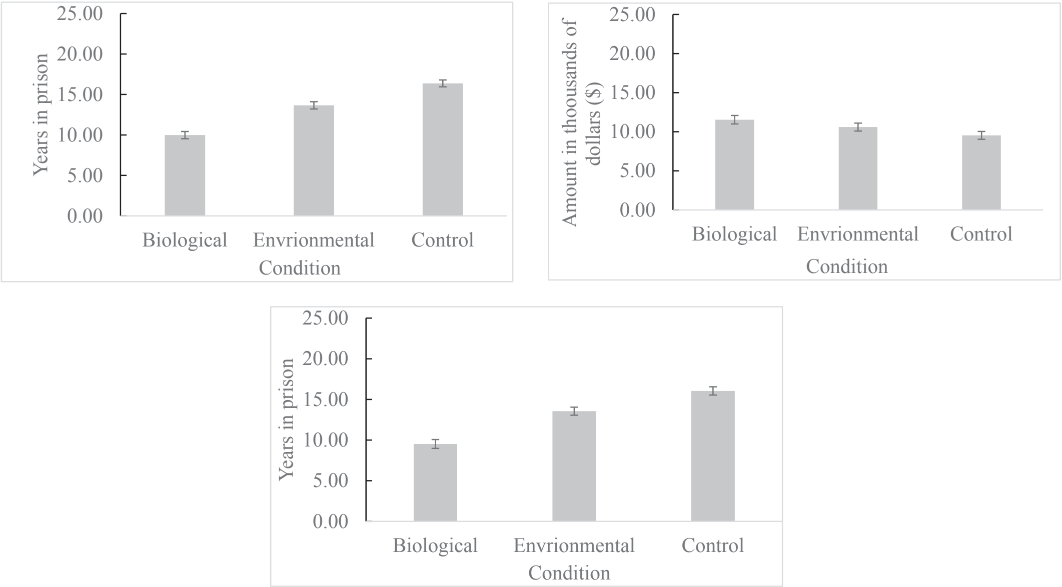


Fig. 2. Decisions by explanation type for punishment (Studies 1 and 2; prison length) and reward (Study 2; reward amount).

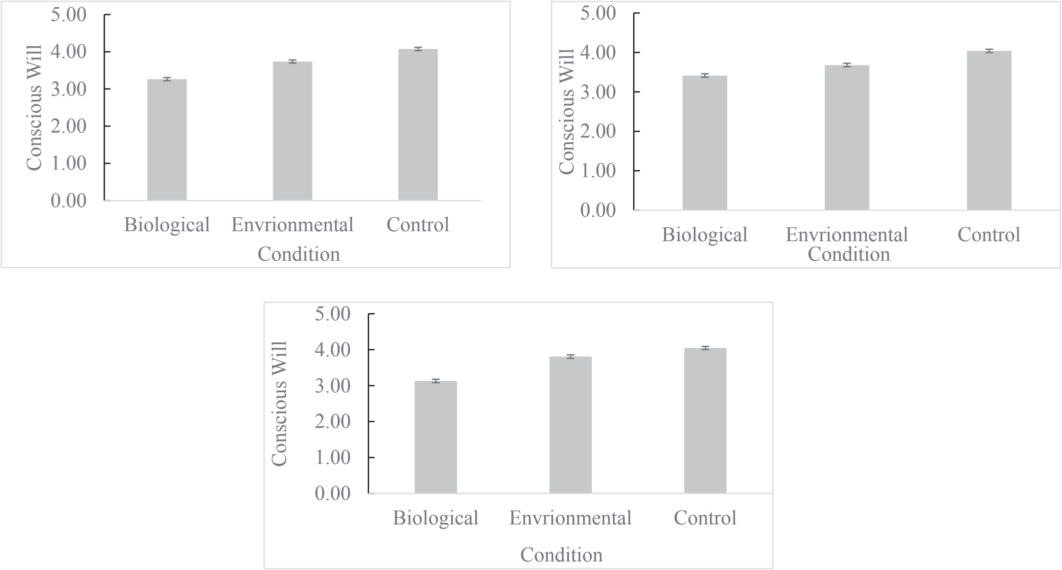
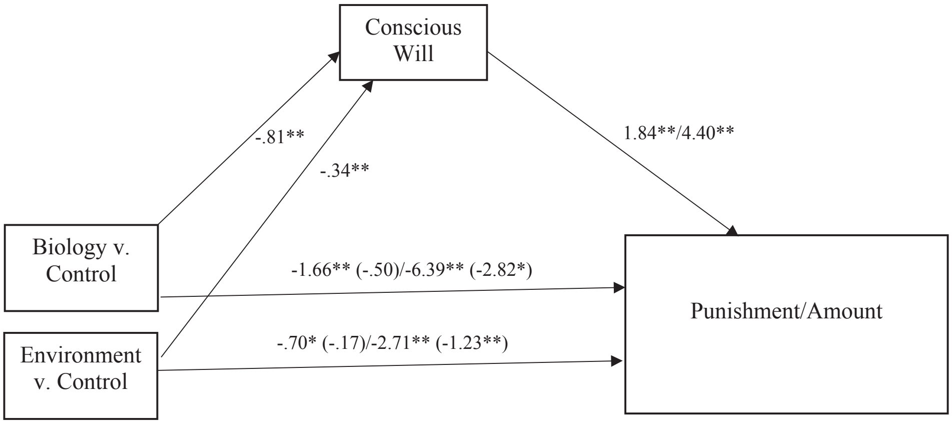


Fig. 3. Conscious will by explanation type for punishment (Studies 1 and 2) and reward (Study 2).

followed a similar stepwise pattern as seen for convictions and prison sentences: participants in the biological condition perceived William as having the least agency, followed by the environmental condition, followed by the control condition. Specifically, reading about biological considerations led to lower perceptions of free will relative to the environmental condition (and especially so compared to the control condition).

3.2. Mediation analysis

To formally test the relationship between explanations of behavior and reductions in convictions and prison sentence length, we ran two mediation models, one for each outcome (conviction, prison sentence length). Biological circumstances predicted both conscious will and each outcome (compared to environmental circumstances and the control condition; see above) and conscious will predicted each outcome ($rs > 0.48$), thus justifying the use of mediation (Baron & Kenny, 1986). We used Hayes' PROCESS macro (i.e., Model 4 with 5000 bootstrap samples; v 3.1). Explanation type was added as the X variable and specified as an indicator type of predictor (i.e., dummy coded for the effects of biological and environmental explanations, relative to the control condition). Conscious will was added as a mediator. Punishment decision was entered as the dependent variable (yes/no for one analysis and 0–25



Indirect Effect Estimates

Punishment
Biology (v. control): -1.49 [-1.86, -1.20]
Environment (v. control): -.62 [-.87, -.40]

Punishment Amount
Biology (v. control): -3.57 [-4.31, -2.89]
Environment (v. control): -1.48 [-2.02, -1.00]

Mediating effect of conscious will on the relationship between biological circumstances and punishment. Asterisks indicate significant effects (*$p < .05$, **$p < .001$). Punishment decision is on the left side of the slash; punishment amount is on the right side of the slash. Unstandardized effects are reported.

Fig. 4. Mediation model for Study 1.

scale for a second analysis). In multicategorical mediation analyses, two indirect paths (one for biological and one for environmental explanations) through conscious will are produced.

As shown in Fig. 4, both biological and environmental circumstances were associated with lower perceptions of conscious will, which in turn were associated with less punishment across both outcomes. Point estimates of the indirect effects with confidence intervals that exclude zero suggest significant mediation (see bottom of Fig. 4).[4]

4. Study 1 discussion

Study 1 found that, upon learning of biological (e.g. brain tumor) or environmental (e.g. parental abuse) circumstances, support for retributive punishment declined. The effect of these circumstances on convictions was mediated by perceptions of how much conscious will the individual had over the transgression. Among those who chose to convict William, participants recommended less severe prison sentences as perceptions of will declined. Biological factors made a greater impact on participant's judgements compared to the environmental factors. This suggests that knowledge of particular determinants of human behavior that are (perceived to fall) outside the domain of conscious control decreases the desire for retribution. These results are in line with previous research demonstrating that moral responsibility hinges on perceptions of an individual's capacity to do otherwise (Bayles, 1982; Cashmore, 2010; Nichols & Knobe, 2007; Pizarro & Tannenbaum, 2011). We suggest that retributive punishment decisions depend on moral reasoning that weighs constraints on conscious will imposed by biological and environmental circumstances (especially biological).

5. Study 2

If reductions in perceptions of conscious will lead to reductions in the desire to elicit punishment for moral transgressors, how do reductions in perceptions of will affect the desire to reward moral paragons? Study 2 aimed to answer this question by adding three more conditions that reversed the negative act deployed in the vignettes of the first study from taking a life to saving a life. A similar vignette structure was used in order to more directly compare results of Study 2 to Study 1.

[4] A reviewer suggested that biological and environmental explanations may reduce punishment through engendering sympathy and not reducing conscious will. In an exploratory analysis (set forth in our preregistration), we tested both conscious will and feelings of sympathy as joint mediators in both studies. We found that reductions in conscious will still reduced punishment decisions, although sympathy was also a significant mediator of this process. Thus, conscious will was an independent (and larger) predictor of punishment decisions than merely feeling sympathy toward William.

5.1. Method

5.1.1. Participants

One thousand, five hundred and sixty-seven (N = 1567) participants (M_{age} = 36.29, SD = 11.06, 44.6% Female) were recruited from Amazon Mechanical Turk (MTurk). Participants received $0.50 for participating in the study. Racial/ethnic composition of the study was as follows: 76.8% white/Caucasian, 7.0% black/African American, 6.9% Hispanic/Latino, 6.0% Asian, 2.2% multiracial, and 1.1% other ethnicities. An additional 79 participants were excluded for saying they did not take the study seriously when asked directly on two questions in the survey. See the OSF site for pre-registration details regarding the analytic plan and analysis. Again, an a priori analysis based on effect sizes from pilot studies (see supplementary materials; d = 0.55) suggested a recommended sample size of 176 (with 80% power at α = 0.05). However, we wanted to ensure that, given the greater complexity of Study 2 (six conditions), we could reliably detect an effect *and* estimate its effect size with more precision. Thus, we chose a smaller effect size (d = 0.20), so we aimed to collect at least 1289 participants.

5.1.2. Procedure and measures

5.1.2.1. Scenario and reward. All participants read a description of William and his action. Participants were randomly assigned to one of six conditions (i.e., a 3 (biological/environmental/control) × 2 (valence) design). The first factor (3 levels) varied the information that arose in the neurological or psychological evaluation after the incident. Either William was found to have a brain tumor, history of abuse, or no apparent explanation of behavior. The second factor (2 levels) varied the valence of the act William committed while at work. William either killed a coworker (as in Study 1) or saved a coworker from a fire. Participants were informed that William had sustained minor injuries during the events and would be evaluated at the hospital.

In the two negative act conditions, William had either a brain tumor or a history of abuse that, according to a hospital physician, affects people's ability to regulate their motivations and emotions, which can lead to outbursts of aggression. In the two positive act conditions, William had either a brain tumor or a history of abuse that, according to a hospital physician, affects people's ability to regulate their motivations and risk taking, which can lead to outbursts or helping others at a high cost to oneself.

The reward conditions acted analogously to the punishment conditions in Study 1. In the punishment conditions, William received treatment and future punishment would not be a deterrent of similar acts. In the context of positive acts, an equivalent reward condition would involve alleviating his underlying condition and any additional reward would not encourage similar positive acts. Thus, we constructed a “positive” condition in which William, with various biological or environmental circumstances, would be rehabilitated and was being offered an award in private (such that rewarding him while not in the presence of others would likely not encourage future good acts of others).

Specifically, participants were randomly assigned to one of six conditions. They were as follows:

(1–3) three conditions in which William kills a coworker in a shooting that are described in Study 1.
(4) William rescuing a coworker; possessing a brain tumor in areas that, according to the hospital physician; affect people’s ability to regulate motivations and risk taking which can lead to outbursts of helping others at a risk to oneself; undergoing a 100% effective treatment to prevent future effects of the brain tumor; and being up for a private award.
(5) William rescuing a coworker; being the survivor of parental abuse, which according to a hospital physician, affects people’s ability to regulate their motivations and risk taking which can lead to outbursts of helping others at a high cost to oneself; undergoing a 100% effective treatment to prevent future effects of the abuse; and being up for a private award.
(6) William rescuing a coworker, with no explanations of behavior, undergoing a 100% effective treatment for his minor injuries, and being up for a private award.

In the punishment conditions, participants were asked whether they would convict and give William a prison sentence on a sliding scale from 0 to 25 (representing up to 25 years). In the reward conditions, participants were asked whether they would award and give William a cash reward on a sliding scale from 0 to 25 (representing up to $25,000), respectively. We used this scale in particular to give a common metric to compare punishment and reward in the context of a 3 × 2 ANOVA.

Lastly, participants were asked about William’s conscious will by using the questions from the first study (α = 0.88).

6. Results

6.1. Biological and environmental influences on conviction/reward and conscious will

See the OSF site for the pre-registered data analytic plan.

6.1.1. Conviction/reward

A logistic regression was constructed predicting the decision (0 = no conviction/reward; 1 = conviction/reward) from biological condition (dummy code), environment condition (dummy code), valence (−1 = negative act; 1 = positive act), a biological condition × valence interaction, and an environmental condition × valence interaction. Due to the dummy coded nature of the variables, these two-way interactions will examine if the effects of biological/environmental conditions (v. control) affect decision making more so when William is being punished or rewarded.

The results of this analysis can be seen in Table 1. Participants in the biological conditions were less likely to grant a conviction/

Table 1
Logistic regression predicting decisions in Study 2.

	b	SE	Wald	p	Exp (b)	95% Confidence Interval	
						LB	UB
Intercept	2.96	0.20	210.73	< .001	19.29		
Biological	−0.89	0.27	10.83	.001	0.41	0.24	0.70
Environmental	−0.02	0.30	0.01	.94	0.98	0.54	1.76
Valence	−0.26	0.20	1.61	.20	0.77	0.52	1.15
Biological × Valence	1.39	0.27	26.75	< .001	4.02	2.37	6.82
Environmental × Valence	0.92	0.30	9.38	.002	2.51	1.39	4.51

Note. $\chi^2(5) = 118.86$, $p < .001$, $R^2 = 0.16$.

reward relative to control. The effects of the environmental condition and valence were not significant. However, these effects are conceptually confusing because they conflate explanation type (for the effect of valence) and valence (for the effects of explanation type). For example, the main effect of biological condition is predicting both punishment and rewards in one analysis, which can be confusing. Of interest are the ways in which explanation and valence conditions *interact* to predict decision making. Indeed, the two two-way interactions were significant, suggesting the effects of explanation type on decision making vary with respect to whether participants were punishing or rewarding William.

As seen in middle and right panels of Fig. 1, explanation type was a significant predictor of punishment ($\chi^2(2) = 70.79, p = .001$) but not reward ($\chi^2(2) = 4.40$, $p = .11$). As in Study 1, there was a monotonic increase in conviction across conditions, such that biological < environmental < control. Specifically, among those who read about William having a brain tumor, a smaller percentage convicted him (71.9%) compared to those who read about William enduring parental abuse (90.7%), or the control condition (96.2%). Like Study 1, all the conditions were significantly different than one another such that participants in the environmental condition had significantly lower conviction rates compared to the control, but this difference was small. The reward decisions did not differ across the conditions.

ANOVAs were used to predict prison sentence length/reward amount and conscious will from the experimental conditions.

6.1.2. Sentence/amount

A 3 (explanation type) × 2 (valence) ANOVA predicting sentence/reward amount revealed significant differences among the conditions, $F(5, 1561) = 24.70, p < .001, \eta^2 = 0.07$. The main effect of explanation type ($F(2, 1561) = 9.99, p < .001, \eta^2 = 0.01$) and valence ($F(1, 1561) = 34.62$, $p < .001$, $\eta^2 = 0.02$) were each significant. Briefly, participants punished more than they rewarded and sentencing/rewards were lowest in the biological condition. Again, these main effects collapse across the other dimension respectively (i.e., the main effect of explanation type ignores whether participants were punishing or rewarding). Of interest was the hypothesized two-way explanation × valence interaction, which was also significant, $F(2, 1561) = 34.39$, $p < .001$, $\eta^2 = 0.04$.

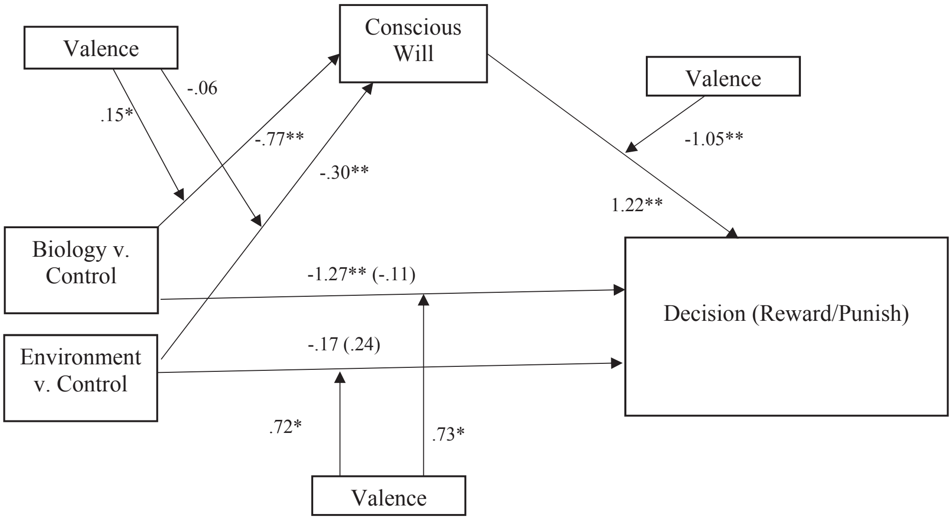
As seen in the middle and right panels of Fig. 2, explanation type predicted sentencing in a straightforward way, $F(2, 775) = 40.57$, $p < .001$. Punishment followed a monotonic pattern in which sentencing was lowest among participants in the biological condition, followed by the environmental condition, followed by the control condition, replicating Study 1. All of these conditions were significantly different than one another ($p < .001$). Although the overall F statistic was significant for reward amount, explanation type inconsistently predicted reward amount, $F(2, 786) = 3.79$, $p = .02$. Nearly all of the conditions were not significantly different than one another ($ps > 0.31$). The one exception was a difference between the biological condition and the control condition, such that rewards amounts were lower in the biological condition compared to the control condition. But this difference was relatively small, $d = 0.24$, $p = .02$.

6.1.3. Conscious will

A 3 (explanation type) × 2 (valence) ANOVA predicting conscious will revealed significant differences among the conditions, $F(5, 1554) = 69.67, p < .001, \eta^2 = 0.18$. The main effect of explanation type ($F(2, 1554) = 161.34, p < .001, \eta^2 = 0.17$) was significant but the effect of valence was not significant, $F(1, 1554) = 1.77$, $p = .18$, $\eta^2 = 0.001$). Participants in the biological condition bestowed lower free will on William, followed by the environmental condition, followed by the control condition (recreating the "stepwise" pattern found with other variables and in Study 1; see middle and right panels of Fig. 3). These differences between explanation types were significant, $ps < 0.001$. The two-way explanation × valence interaction was also significant, $F(2, 1554) = 11.88$, $p < .001$, $\eta^2 = 0.02$. This significant two-way interaction was primarily driven by differences in the biological condition in the reward v. punishment condition ($d = 0.40, p < .001$), such that participants thought William's brain tumor did not reduce conscious will as much when he saved a coworker than when murdered a coworker. The other conditions do not differ from each other across valence conditions ($ps > 0.16$).

6.2. Mediation analyses

We conducted a multicategorical moderated mediation model (model 59) using Hayes' (2016) PROCESS macro (v 3.1).



Moderated Mediation Estimates
Biology (v. control): 1.98 [1.40, 2.76]
Environment (v. control): .48 [.11, .93]

Indirect Effect Estimates
Punishment
Biology (v. control): -2.09 [-2.69, -1.65]
Environment (v. control): -.54 [-.89, -.25]

Reward
Biology (v. control): -.11 [-.50, .38]
Environment (v. control): -.06 [-.29, .22]

Mediating effect of conscious will on the relationship between biological circumstances and punishment. Asterisks indicate significant effects (*$p < .05$, **$p < .001$). Unstandardized effects are reported.

Fig. 5. Moderated mediation model for Study 2 predicting reward/punishment decisions.

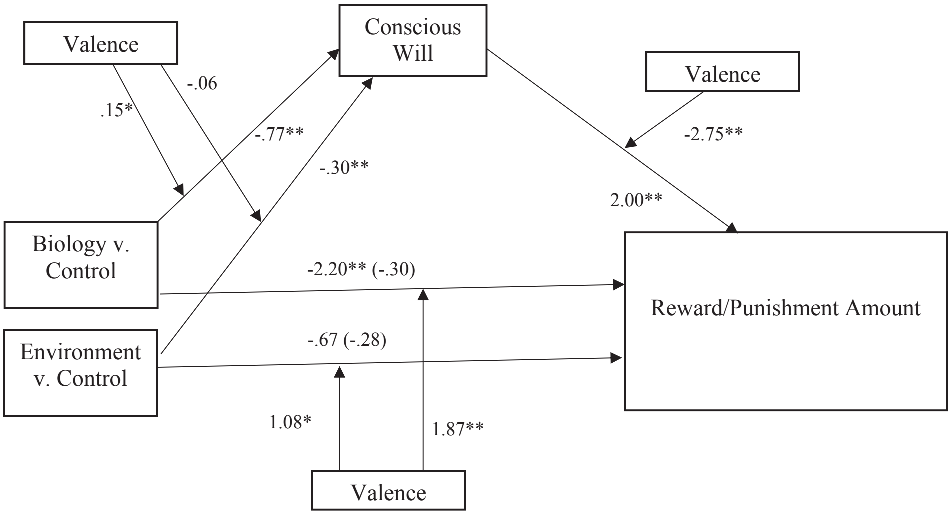
Explanation type was added as the X variable and specified as an indicator type of predictor (i.e., dummy coded for the effects of biological and environmental explanations, relative to the control condition). Conscious will was added as a mediator. Decision (reward/punish) was entered as the dependent variable (yes/no for one analysis and 0–25 scale for a second analysis). Valence (−1: negative; 1: positive) was entered as the moderator of the a, b, and c paths. This moderation model tested if explanation type affects decisions through conscious will differently depending on whether participants were rewarding versus punishing William. In multicategorical mediation analyses, two indirect paths (one for biological and one for environmental explanations) through agency are produced.

To ease interpretation, we produced multiple figures. In Figs. 5 (decision: punish/reward) and 6 (prison sentence/reward amount), we present the full moderated mediation analysis. However, because the main effects in the context of moderation mediation again conflate across valence, they are difficult to interpret. Thus, in Fig. 7, we produce simpler mediation figures for punishment (left) and reward (right) separately. Put briefly, the hypothesized moderated mediation was supported for both reward/punishment decisions (Biology: 1.98 [1.40, 2.76]; Environment: 0.48 [0.11, 0.93]) and reward/punishment amount (Biology: 4.83 [3.77, 5.88]; Environment: 1.41 [0.70, 2.13]). Thus, explanation type affected conscious will, which then affected decisions, but this process differed depending on whether William was being punished or rewarded.

Specifically, we replicated the findings from Study 1, such that conscious will mediated the effects of explanation type on punishment decisions/amount (see left panel of Fig. 7 for estimates and indirect effects) but not reward decisions/amount (see right panel of Fig. 7 for estimates and indirect effects). In nearly every case, valence moderated the effects, such that explanation types (and conscious will) were associated with lower (higher) punishment. Thus, lowered perceptions of will affected punishment decisions but not reward decisions.

7. Study 2 small discussion

Consistent with Study 1, Study 2 found that biological and environmental circumstances diminished perceptions of conscious will. Once again, biological circumstances demonstrated larger effects on conscious will compared to environmental circumstances. In



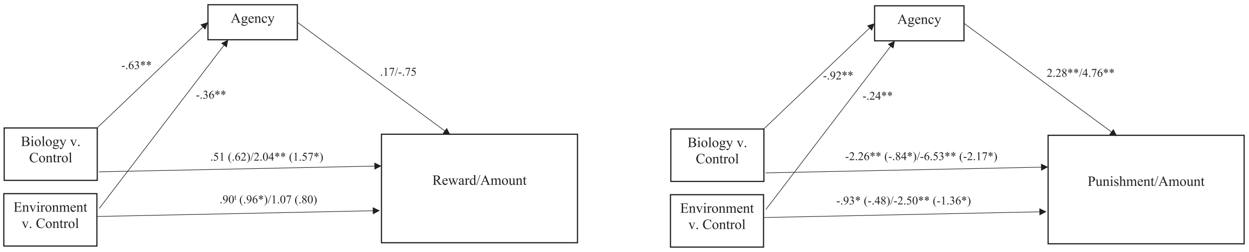
Moderated Mediation Estimates
Biology (v. control): 4.83 [3.77, 5.88]
Environment (v. control): 1.41 [.70, 2.13]

Indirect Effect Estimates
Punishment
Biology (v. control): -4.36 [-5.29, -3.48]
Environment (v. control): -1.14 [-1.78, -.51]

Reward
Biology (v. control): .47 [-.11, 1.02]
Environment (v. control): .27 [-.06, .64]

Mediating effect of conscious will on the relationship between biological circumstances and punishment. Asterisks indicate significant effects (*p < .05, **p < .001). Unstandardized effects are reported.

Fig. 6. Moderated mediation model for Study 2 predicting reward/punishment amount.



Indirect Effect Estimates
Reward
Biology (v. control): -.11 [-.50, .38]
Environment (v. control): -.06 [-.29, .22]

Reward Amount
Biology (v. control): .47 [-.11, 1.02]
Environment (v. control): .27 [-.06, .64]

Mediating effect of agency on the relationship between biological circumstances and punishment. Asterisks indicate significant effects (^{t} p < .10, *p < .05, **p < .001). Reward decision is on the left side of the slash; reward amount is on the right side of the slash.

Indirect Effect Estimates
Punishment
Biology (v. control): -2.09 [-2.69, -1.65]
Environment (v. control): -.54 [-.89, -.25]

Punishment Amount
Biology (v. control): -4.36 [-5.29, -3.48]
Environment (v. control): -1.14 [-1.78, -.51]

Mediating effect of agency on the relationship between biological circumstances and punishment. Asterisks indicate significant effects (^{t} p < .10, *p < .05, **p < .001). Punishment decision is on the left side of the slash; punishment amount is on the right side of the slash.

Fig. 7. Mediation models decomposed into punishment and reward.

addition, decreased perceptions of conscious will from behavioral explanations subsequently reduced punishment judgements, as in Study 1. Despite this, subsequent reward judgements did not vary with respect to reductions in conscious will from biological or environmental explanations. In other words, our results suggest that when evaluating a heroic act, the scope of an actor's conscious will does not appear to matter for rewarding them. Instead, individuals who commit heroic actions are rewarded uniformly, regardless of how the behavior originated in the individual.

8. General discussion

Two studies investigated the consequences of reductions in perceptions of conscious will from biological and environmental circumstances. A reduced sense of a criminal's will diminished support for punishment (Studies 1 and 2). These results illustrate the importance of evaluating the spectrum of conscious will of a trespasser in the context of a crime. However, a similar reduced sense of conscious will had no effect on support for the reward of a hero in the context of a virtuous act (Study 2). Instead, judgements of reward were given invariably, regardless of the origin of the behavior. Across both studies, biological considerations consistently demonstrated significantly higher reductions in perceptions of conscious will compared to environmental considerations.

Within Studies 1 and 2, participants consistently elected to sentence William to time in prison, even though they were informed the potential social benefits were already accomplished (e.g., the rehabilitation reduced the chance of re-offense to zero and others' actions would likely not be affected). Why did participants still allocate prison sentences even though he was viewed as having little control over his actions and that punishment would offer no further rehabilitative benefits? Previous research has suggested that even though consequentialist benefits are people's stated preference (i.e., to prevent future negative actions), the primary motivation for punishment is almost entirely retributive (i.e., feeling as though William deserves to be punished; Carlsmith, Darley, & Robinson, 2002). The widespread desire for retribution and third-party punishment may actually be a product of a means to enforce cooperation and order among individuals within society (Boyd, Gintis, Bowles, & Richerson, 2003; Cushman, 2013; Daly & Wilson, 1988; Fehr & Fischbacher, 2004). This is because cooperation demands the overriding of self-interested behavior among individuals in a group (Henrich et al., 2006). An individual's most efficient decision in group activities is taking a benefit without providing a cost (including committing a murder without enduring consequences). However, if such a strategy becomes widespread, the group and society will collapse. In this way, punishment generally ensures behavior is steered towards group benefits and away from self-interests (Fehr & Gachter, 2000; Fehr, Gächter, & Kirchsteiger, 1997; McCullough, Kurzban, & Tabak, 2010). Retributive punishment then serves as a mechanism to hold oneself and others morally responsible to promote cohesion. Such a disposition towards retribution is possible due to a persistent and prevalent belief in free conscious will (Greene & Cohen, 2004; Monroe & Malle, 2010; Sarkissian et al., 2010; Wegner, 2002). An extreme belief in free conscious will entails that one is morally responsible for *every* behavior that occurs. However, this position is not the abstraction most people hold, and most would likely not make this claim. Instead, the lay belief of conscious will recognizes that mental states and behavior can be subject to mitigation or explanations. Because of this, endorsement for retribution varies depending on awareness of an individual's particular circumstances (e.g., biological and environmental limitations on conscious will).

8.1. The role of agency in reward and punishment – Free will's asymmetry in punishment and reward

The importance of conscious will in the context of moral transgressions and punishment raises the question of why conscious will is not salient in judgements of reward. That is, why is the scope of conscious will involved in evaluations of criminal acts but not virtuous acts? After all, Study 2 contextualized the virtuous act as not arising entirely from an individual's own will (instead, resulting from a biological or environmental factor). Yet, subsequent reductions in perceived conscious will from biological and environmental circumstances did not translate to reductions in reward decisions. One possible explanation is that punishment requires more justification than reward, particularly that an agent consciously committed a negative act. The question of why this asymmetry between positive and negative acts exists is discussed next.

Reward and encouragement of future prosocial acts in Study 2 was high regardless of the perceived origin of the positive act. If William's conscious will initiated the virtuous action, then an award may encourage similar future behavior through encouraging his conscious decision process. If, however, the behavior resulted from biological and environmental causes that were then removed (see Method), reward may still encourage similar future behavior that is not influenced by these circumstances (e.g., reward reinforces doing heroic things). That is, we likely reward those who commit virtuous acts even if they did not have complete conscious control over their actions, possibly because people want to encourage others to engage in virtuous acts. In both cases, reward and its function remain the same. Indeed, rewarding positive acts encourages similar prosocial future behavior (Aknin, Van de Vondervoort, & Hamlin, 2018; Waugh, Brownell, & Pollock, 2015). This may explain why it is not necessarily an imperative to distinguish between causes of behavior and one's will when encouraging future behavior in terms of reward. Either way, people want others who commit good deeds to continue doing so.

However, discouragement of future criminal acts in Studies 1 and 2 varied depending on the perceived origin of the negative act. If William's conscious will initiated the criminal action, punishing him may discourage similar future behavior so he doesn't do bad things again. If, however, the behavior resulted from biological and environmental causes which were then removed, punishment may once again discourage similar future behavior (e.g., punishment still serves to dissuade William from doing bad things without the circumstances present). Importantly, however, this may be viewed as unjust (e.g., being punished for a negative action that one did not consciously choose). Indeed, when a negative act was perceived to be caused by biological and environmental circumstances that were retroactively removed in Studies 1 and 2, punishment rates dropped. This might be because the cause of the behavior is explained as occurring outside of one's conscious will, and the locus of discouragement to act on is no longer present (e.g., punishment occurs to deter a cause that no longer exists). Punishment to dissuade future acts becomes increasingly similar to penalizing an innocent individual, as the current individual is not viewed to possess the original cause of the behavior (e.g., the biological/environmental factor) that is to be discouraged, reformed, or protected against. In other words, if the perceived cause of a negative act is removed entirely, is it appropriate to punish that person to the same degree as someone who consciously caused their action? It follows that punishment decisions seem to change depending on the cause. In contrast to reward decisions, this may explain why it is important to distinguish between causes of behavior and to examine one's level of conscious control when discouraging future behavior in terms of punishment. People may not want to punish a person who is not viewed to be the direct, conscious cause of a negative act.

In line with this idea, in most parts of the world, punishment given to individuals without taking into account their agency and will is considered inhumane and unethical (Alexander, 1983; Smilansky, 1990). Even in situations in which people are punished for being negligent, the fact that they had the choice to direct their attention to a problem assumes some degree of conscious will. However, rewarding an individual for a positive act without evaluating their agency and will does not ring any alarms and is not considered inhumane or unethical.

Much of society is organized around the fact that we hold people accountable for negative acts, whether it be to prevent their future occurrence (both by the perpetrators and others) or to seek retribution for that particular negative act. The equivalent reasoning for virtuous acts is not as straightforward, at least according to our studies. Rates of reward were relatively high across conditions, suggesting that detriments in conscious will do not drive reward decisions. But what does drive reward decisions? Whether analogous processes that explain both the discouragement of antisocial behavior and the encouragement of prosocial behavior exist is unclear. We hope that future researchers can help develop models that explain how one's will intersects with judgement and decision making for a variety of actions.

Future work should also continue to analyze the relationship between conscious will and reward. For example, if conscious will attributions allow individuals to hold each other responsible to limit antisocial behavior, why are they not also deployed to encourage prosocial behavior? Is an actor who was manipulated by another into committing a good deed (e.g., a teacher instructing a child to be nice to a peer) as deserving of reward as someone who seemingly willed their good deed (e.g., a child who is spontaneously nice to their peer)?

8.2. Judgements of free will

In the present study, potential biological explanations for behavior led to greater reductions in perceptions of conscious will relative to environmental explanations. This is in line with previous research that demonstrates a preference for attributing a behavior to individual dispositions (i.e. genetics, internal factors) compared to social/contextual factors (Dar-Nimrod & Heine, 2011). Biological influences could be perceived as being the more proximate and deterministic cause of a behavior. Nevertheless, both factors consistently reduced perceptions of conscious will (Studies 1 and 2). The exact extent to which a biological or environmental consideration actually increases the probability of an action in real life cannot currently be known, particularly to lay people. Overestimates of how biological circumstances reduce agency may lead to discounting the role that other factors have on behavior, including both additional biological circumstances and environmental circumstances. This discrepancy may lead to errors in judgements of how a behavior occurred, and therefore affect subsequent consequences or interpretations of that behavior.

It is also worth acknowledging that the relationship between perceptions of free will and judgements of punishment is likely bidirectional. Although we have seen how perceptions of compromised will reduce the desire to punish, an impulse to punish may also culminate in increased perceptions of will. In other words, we may initially desire to punish someone and then subsequently make judgements about their conscious will post hoc. Indeed, Clark et al. (2014) found that escalating the motivation to impose punishment (e.g., such as towards a cheater on a test) increased perceptions that the perpetrator had the capacity to consciously will their behavior. This is likely due to affective motivated reasoning, in which judgements are guided primarily by emotional intuition rather than more cognitive, deliberative processes (Haidt, 2001). In this case, an increased initial urge to punish a wrongdoing led to post hoc justifications toward the trespasser's conscious will. An increased motivation to punish may therefore lead to greater perceptions of conscious will even in the context of individuals experiencing biological and environmental circumstances that ostensibly put restrictions on their will.

Further, one may worry that, because the concepts of conscious will and moral responsibility are so connected, a diminished view of conscious will might lead to an increased prevalence of antisocial and self-interested behavior. Indeed, some preliminary research has indicated that when free will beliefs are experimentally reduced, people are less helpful to others and may be more aggressive (Baumeister, Masicampo, & DeWall, 2009; but see Embley, Johnson, & Giner-Sorolla, 2015). Thus, there are large societal implications for lay beliefs regarding free will and how these beliefs extend to interpersonal behavior. However, the present studies suggest an additional insight into the consequences of reduced perceptions of conscious will, particularly when making judgements about other people. Together with previous research (Caspar, Vuillaume, Magalhães De Saldanha da Gama, & Cleeremans, 2017; Krueger, Hoffman, Walter, & Grafman, 2013; Shariff et al., 2014), we suggest that the desire to balance the scale of justice declines when perceptions of conscious will are reduced. This does not suggest that responsibility gets thrown out altogether. As previously noted, causal responsibility for a crime remains viable in consequentialist punishment, which does not depend on the belief in free will (Greene & Cohen, 2004; Shariff et al., 2014). In fact, conviction rates were still above 50% in all conditions across Studies 1 and 2, even in the context of biological and environmental circumstances. In addition, the current studies demonstrate that reduced perceptions of free will do not lead to diminished support for reward. Instead, individuals who commit heroic actions are rewarded uniformly, regardless of how the behavior originated in the individual. In other words, in the context of an absence of conscious will, encouragement of prosocial acts remains.

8.3. Limitations

One limitation for interpreting results from the current study is that participants responded to hypothetical vignettes about a moral transgression and biological and environmental circumstances about an individual. Thus, they were not exposed to the emotional intensity of in-person interactions with a moral trespasser (e.g., as a real juror in a real courtroom). It appears likely that such confrontations would raise affective intuitions towards punishment and in turn perceptions of conscious will. Likewise, the prospect of an entirely successful rehabilitation program might have struck many participants as unrealistic, although 66–72% thought it was believable and the results remained after controlling for believability of the stimuli (see Footnote #2). As with any experimental study, researchers must weigh how many details they provide participants with (e.g., more individuating details about William), so that a research question of interest can be isolated. An in-situ experiment would likely introduce many additional sources that might affect attributions of guilt and free will (e.g., race). The design of the study prevented participants from being unduly influenced by additional aspects of the individual's personality and behavior that are unrelated to the moral wrongdoing. Nevertheless, future research should further examine the conditions under which constraints on conscious will affect the desire to punish in more realistic settings.

In addition, the participants in each study were entirely from MTurk, who have limited generalizability to other samples (Henrich, Heine, & Norenzayan, 2010). Future research should investigate the connection between conscious will and punishment both cross-culturally and across socio-economic class (Clark et al., 2014; Kraus, Piff, & Keltner, 2009).

8.4. Conclusion

The current studies investigated the consequences of altering perceptions of conscious will and the desire for punishment and reward through knowledge of biological or environmental circumstances. In punishment decisions, we found that biological circumstances (e.g., brain tumors) diminished judgements of a perpetuator's will to a greater extent than environmental circumstances (e.g., child abuse). In turn, reductions in perceived conscious will led to decreased support for retributive punishment. In the context of reward decisions, deficits in conscious will did not translate to a lower likelihood of reward, suggesting that free will beliefs may not be an integral factor in reward decisions. Because punishment serves as a crucial instrument for human cooperation, understanding variations in the desire for retribution is essential. Whether it is beneficial for society to have a reduced urge to balance the scale of justice when criminals are deemed less consciously in control of their behavior is open for debate. In either case, the results of these studies imply there are societal consequences in response to alterations in public perceptions of an agent's genuine capacity to have done otherwise. Future research is needed to understand the broader implications of the intimate relationship between conscious will and punishment in judgements of moral transgressors and whether conscious will plays a role in judgements of moral paragons.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.concog.2019.102808.

References

Aknin, L. B., Van de Vondervoort, J. W., & Hamlin, J. K. (2018). Positive feelings reward and promote prosocial behavior. *Current Opinion in Psychology, 20*, 55–59.

Alexander, L. (1983). Retributivism and the Inadvertent Punishment of the Innocent. *Law and Philosophy, 2*(2), 233–246.

Alicke, M. (2000). Culpable control and the psychology of blame. *Psychological Bulletin, 126*(4), 556–574.

Aspinwall, L. G., Brown, T. R., & Tabery, J. (2012). The double-edged sword: Does biomechanism increase or decrease judges' sentencing of psychopaths? *Science, 337*(6096), 846–849.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323.

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*(2), 260–268.

Bayles, M. D. (1982). Character, purpose, and criminal responsibility. *Law and Philosophy, 1*(1), 5–20.

Bentham, J. (1986). *An introduction to the principles of morals and legislation: Printed in the year 1780, and now first published.* Legal Classics Library.

Bernet, W., Vnencak-Jones, C. L., Farahany, N., & Montgomery, S. A. (2007). Bad nature, bad nurture, and testimony regarding MAOA and SLC6A4 genotyping at murder trials. *Journal of Forensic Sciences, 52*(6), 1362–1371.

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences, 100*(6), 3531–3535.

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology, 83*(2), 284.

Cashmore, A. R. (2010). The Lucretian swerve: The biological basis of human behavior and the criminal justice system. *Proceedings of the National Academy of Sciences, 107*(10), 4499–4504.

Caspar, E. A., Vuillaume, L., Magalhães De Saldanha da Gama, P. A., & Cleeremans, A. (2017). The influence of (dis) belief in free will on immoral behavior. *Frontiers in Psychology, 8*, 20.

Clark, C. J., Luguri, J. B., Ditto, P. H., Knobe, J., Shariff, A. F., & Baumeister, R. F. (2014). Free to punish: A motivated account of free will belief. *Journal of Personality and Social Psychology, 106*(4), 501.

Clark, C. J., Shniderman, A., Luguri, J. B., Baumeister, R. F., & Ditto, P. H. (2018). Are morally good actions ever free? *Consciousness and Cognition*.

Cooper, J. A., Walsh, A., & Ellis, L. (2010). Is criminology moving toward a paradigm shift? Evidence from a survey of the American Society of Criminology. *Journal of Criminal Justice Education, 21*(3), 332–347.

Cushman, F. (2013). The role of learning in punishment, prosociality, and human. *Cooperation and its Evolution, 333*.

Cusimano, C. J., & Goodwin, G. P. (2019). Folk attributions of control and intentionality over mental states. *Cognitive Science*.

Daly, M., & Wilson, M. (1988). *Homicide.* Transaction Publishers.

Dar-Nimrod, I., & Heine, S. J. (2011). Genetic essentialism: On the deceptive determinism of DNA. *Psychological Bulletin, 137*(5), 800.

Embley, J., Johnson, L. G., & Giner-Sorolla, R. (2015). Replication of Study 1 by Vohs & Schooler (2008, Psychological Science). Replication part of the Reproducibility Project. Retrieved from: https://osf.io/2nf3u/.

Farahany, N. A., & Coleman, J. E. (2006). Genetics and responsibility: To know the criminal from the crime. *Law and Contemporary Problems, 69*(1/2), 115–164.

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences, 8*(4), 185–190.

Fehr, E., & Gachter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*(4), 980–994.

Fehr, E., Gächter, S., & Kirchsteiger, G. (1997). Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica: Journal of the Econometric Society,* 833–860.

Feldman, G., Wong, K. F. E., & Baumeister, R. F. (2016). Bad is freer than good: Positive–negative asymmetry in attributions of free will. *Consciousness and Cognition, 42*, 26–40.

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind.* Cambridge, MA: MIT Press.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117*(1), 21.

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry, 23*(2), 101–124.

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.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814.

Hamlin, J. K., & Baron, A. S. (2014). Agency attribution in infancy: Evidence for a negativity bias. *PloS One, 9*(5), e96112.

Heberlein, A. S., & Adolphs, R. (2004). Impaired spontaneous anthropomorphizing despite intact perception and social knowledge. *Proceedings of the National Academy of Sciences of the United States of America, 101*(19), 7487–7491.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature, 466*(7302), 29.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., ... Lesorogol, C. (2006). Costly punishment across human societies. *Science, 312*(5781), 1767–1770.

Knobe, J. (2003). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology, 16*(2), 309–324.

Kraus, M. W., Piff, P. K., & Keltner, D. (2009). Social class, sense of control, and social explanation. *Journal of Personality and Social Psychology, 97*(6), 992–1004.

Krueger, F., Hoffman, M., Walter, H., & Grafman, J. (2013). An fMRI investigation of the effects of belief in free will on third-party punishment. *Social Cognitive and Affective Neuroscience, 9*(8), 1143–1149.

McCullough, M. E., Kurzban, R., & Tabak, B. A. (2010). Evolved mechanisms for revenge and forgiveness. *Understanding and Reducing Aggression, Violence, and their Consequences,* 221–239.

Meehl, P. E. (1977). Specific etiology and other forms of strong influence: Some quantitative meanings. *The Journal of Medicine and Philosophy, 2*(1), 33–53.

Monroe, A. E., & Malle, B. F. (2010). From uncaused will to conscious choice: The need to study, not speculate about people's folk concept of free will. *Review of Philosophy and Psychology, 1*(2), 211–224.

Morewedge, C. K. (2009). Negativity bias in attribution of external agency. *Journal of Experimental Psychology: General, 138*(4), 535.

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

Nichols, S. (2004). The folk psychology of free will: Fits and starts. *Mind & Language, 19*(5), 473–502.

Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous, 41*(4), 663–685.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231.

Pizarro, D. A., & Tannenbaum, D. (2011). Bringing character back: How the motivation to evaluate character influences judgements of moral blame. *The Social Psychology of Morality: Exploring the Causes of Good and Evil,* 91–108.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515–526.

Provencher, H., & Fincham, F. D. (2000). Attributions of causality, responsibility, and blame for positive and negative symptom behaviors in caregivers of persons with schizophrenia. *Psychological Medicine, 30*, 899–910.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.). *Advances in experimental social psychology* (pp. 173–220). New York: Academic Press.

Sarkissian, H., Chatterjee, A., De Brigard, F., Knobe, J., Nichols, S., & Sirker, S. (2010). Is belief in free will a cultural universal? *Mind & Language, 25*(3), 346–358.

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*(8), 1563–1570.

Smilansky, S. (1990). Utilitarianism and the'punishment'of the innocent: The general problem. *Analysis, 50*(4), 256–261.

Spinoza, B. (1985). The Collected Works of Spinoza. Vol. 1. Trans. E. Curley. Princeton: Princeton University Press.

Steinberg, L., & Scott, E. S. (2003). Less guilty by reason of adolescence: Developmental immaturity, diminished responsibility, and the juvenile death penalty. *American Psychologist, 58*(12), 1009.

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin, 110*(1), 67.

Walsh, A. (2010). *Biology and criminology: The biosocial synthesis.* Routledge.

Waugh, W., Brownell, C., & Pollock, B. (2015). Early socialization of prosocial behavior: Patterns in parents' encouragement of toddlers' helping in an everyday household task. *Infant Behavior and Development, 39*, 1–10.

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