2-2018

Crime, Punishment, and Causation: The Effect of Etiological Information on the Perception of Moral Agency

Paul J. Litton

Philip Robbins

Crime, Punishment, and Causation: The Effect of Etiological Information on the Perception of Moral Agency

Philip Robbins and Paul Litton
University of Missouri

Moral judgments about a situation are profoundly shaped by the perception of individuals in that situation as either moral agents or moral patients (Gray & Wegner, 2009; Gray, Young, & Waytz, 2012). Specifically, the more we see someone as a moral agent, the less we see them as a moral patient, and vice versa. As a result, casting the perpetrator of a transgression as a victim tends to have the effect of making them seem less blameworthy (Gray & Wegner, 2011). Based on this theoretical framework, we predicted that criminal offenders with a mental disorder that predisposes them to antisocial behavior would be judged more negatively when the disorder is described as having a genetic origin than when it is described as environmentally caused, as in the case of childhood abuse or accident. Further, we predicted that some environmental explanations would mitigate attributions of blame more than others, namely, that offenders whose disorder was caused by childhood abuse (intentional harm) would be seen as less blameworthy than offenders whose disorder is caused by an unfortunate accident (unintentional harm). Results from two vignette-based studies designed to test these predictions, conducted with participants recruited from Amazon Mechanical Turk ($N = 244$ and $N = 387$, respectively), confirmed the first prediction but not the second. Implications of this research for three areas—the psychology of moral judgment, philosophical debates about moral responsibility and determinism, and the practice of the law—are discussed in the sequel.

Keywords: moral typecasting, blame, punishment, responsibility, causation

On July 5, 1978, Robert Alton Harris and his brother spotted two teenage boys eating in the parking lot of a fast food restaurant. Harris forced them into their car at gunpoint and ordered one of them to drive. After stopping and ordering them out of the car, Harris shot one boy in the back, then chased down the other boy and shot him several times. Upon returning to the car, Harris found his first victim still alive and shot him in the head. Afterward, Harris and his brother finished eating the hamburgers that the boys had bought, and then used their car to rob a bank. Arrested and charged with multiple counts of first-degree murder, Harris was found guilty at trial and sentenced to death.

Harris' monstrous behavior might be causally explained on a variety of levels. At least in principle we could explain his conduct and traits physiologically, in terms of abnormalities of brain chemistry and structure. This would be a proximal causal explanation. Facts about his genetic profile could also contribute to a more

distal causal story at the physiological level, explaining why his brain failed to develop normally. Alternatively, we might explain his wrongdoing in environmental terms, pointing to the horrific abuse and neglect that Harris suffered as a child. Indeed, expert witnesses described his childhood as a "prisoner-of-war-camp" and his experiences as "a 9 on a Richter scale for traumatic stress" (Gross, 1990). A third possibility would be to explain Harris's behavior in terms of an interaction between genetic and environmental factors. Individuals with the so-called warrior gene (MAOA-L) who have been abused as children, for example, appear to be predisposed to violent behavior, and Harris may have been such an individual (Meyer-Lindenberg et al., 2006; Tiihonen et al., 2015).

Are physiological or environmental explanations relevant to determining Harris's culpability for his crimes? Aside from this puzzling philosophical question, it is important to know whether people generally find such explanations relevant to blame and punishment. Accordingly, empirical researchers have investigated the extent to which such causal explanations influence ordinary intuitions about appropriate punishment. With scientific knowledge advancing with respect to the causes of antisocial conduct, it should be helpful to lawyers and lawmakers to know the extent to which such evidence may affect judges and juries. Empirical research into the effect of causal explanations on judgments of blameworthiness, moral responsibility, and appropriate punishment could have practical importance for criminal lawyers as well as for courts assessing the obligations of counsel.

Such research is relevant to philosophical inquiry as well. Philosophers working on issues surrounding free will and moral responsibility appeal to ordinary intuitions to support their argu-

ments. Incompatibilists, for example, describe cases in which a hypothetical agent would not be deemed morally responsible by ordinary intuitions, and then argue that the best explanation for the intuition is that the agent's conduct was causally determined. Imagine for the moment that ordinary intuitions find genetic explanations of an agent's criminal behavior to undermine their moral responsibility. A natural interpretation of this hypothetical intuition is that genetic explanations mitigate because the offender's genes, which are outside his control, caused his violence. If that interpretation is correct, then incompatibilists can make inroads with what Moore (1997) calls the "beachhead challenge." The incompatibilist points to an excuse we accept within our moral practices and shows that its rationale is based on causation by events outside an agent's control. If we accept one excuse (the argument's "beachhead") based on causation by factors outside an agent's control, and if determinism is true, then the excuse should generalize to eliminate responsibility entirely—because if determinism is true, all conduct is caused by factors outside our control. If we find offenders less than fully responsible when their behavior is given a genetic explanation, then the incompatibilist can latch onto that beachhead; at the very least, the argument would go, we find causation to diminish an agent's responsibility. To deny the beachhead, the compatibilist must either deny that a genetic explanation can diminish responsibility or provide a compatibilist rationale for the intuition that it can. The same point applies to other kinds of causal explanation, whether at the physiological or environmental level, which might influence our intuitions about responsibility.

To proceed along this path, however, we need to have more than an armchair sense of the candidates for incompatibilist beachheads. Perhaps some kinds of causal explanation affect our intuitions about responsibility but others do not. For example, if intuitions about responsibility are unaffected by genetic explanations of wrongdoing, then compatibilists need not worry about explaining such cases; but if intuitions find environmental causes of wrongdoing to mitigate responsibility, then compatibilists will need to either offer a compatibilist-friendly justification for those intuitions or explain them away. To explore these possibilities, though, we must first systematically investigate what ordinary intuitions say about such cases.

A natural starting point for this investigation is empirical research on the general structure of moral cognition. Of particular relevance to our project is the Theory of Dyadic Morality (TDM), which posits a single cognitive template underlying all moral judgments (Gray & Wegner, 2009, 2011; Gray, Young, & Waytz, 2012; Schein & Gray, 2017). According to TDM, moral judgments about a situation are profoundly shaped by the perception of individuals in that situation as either moral agents or moral patients. By definition, a moral agent has the capacity to perform morally good or bad actions, whereas a moral patient has the capacity to be on the receiving end of such actions. The conceptual dichotomy between agency and patiency is governed by the principle of "moral typecasting": The more we see someone as a moral agent, the less we see them as a moral patient, and vice versa. In other words, moral agency and moral patiency are antithetical roles, and moral actors tend to be cast in one role to the exclusion of the other, even across contexts. For example, casting the perpetrator of a transgression as a victim of harm tends to have the

effect of making them seem less blameworthy (Gray & Wegner, 2011).

In the present context, the significance of TDM as an account of moral cognition is largely because of the fact that it generates clear predictions about how etiological information will influence the way people think about criminal behavior. According to the theory, we should expect that criminal offenders with a mental disorder that predisposes them to violent antisocial behavior will be judged more negatively when the disorder is described as having a genetic origin than when it is described as environmentally caused, as in the case of childhood abuse or accident. The basis of this prediction is as follows. When the disorder has an environmental origin, there is a preexisting person who has suffered harm; hence, the perception of their moral patiency should be heightened, and the perception of their moral agency attenuated, by the addition of etiological information. When the disorder has a genetic origin, by contrast, there is no preexisting person to whom harm has been done (because no person exists before the determination of their genetic profile), so the perception of their moral agency should be unaffected by the receipt of information about the cause of their pathology. In other words, offenders whose disorder arises from environmental causes should be seen as less blameworthy than offenders whose disorder is caused by bad genes, because the former will be seen as victims but the latter will not. Moreover, genetic explanations of psychopathology, insofar as they do not implicate personal harm or victimhood, should not affect the perception of moral agency.

Most of the evidence gathered to date concerning the effects of causal explanations on moral judgment in the context of criminal law is consistent with the account sketched above. That said, some caveats are in order, especially with respect to the issue of whether environmental explanations that clearly implicate victimhood, such as a history of childhood abuse, have a mitigating effect.

Consider first the issue of whether genetic explanations of criminal behavior mitigate judgments of blame, punishment, and responsibility. One study, using state trial court judges as participants, investigated whether a neurogenetic explanation of an offender's psychopathy would affect judgments of responsibility and punishment (Aspinwall, Brown, & Tabery, 2012). All participants read a vignette involving an aggravated battery by an offender given a diagnosis of psychopathy by a psychiatrist providing expert testimony on the case. Half the participants then read additional expert testimony by a neurobiologist, to the effect that the offender's psychopathy resulted from a genetic defect linked to abnormal development of brain structures involved in emotion processing. The addition of this explanation significantly reduced both the extent to which psychopathy was rated as aggravating and the severity of sentencing, contrary to what TDM predicts. It is possible, however, that the mitigating effect of the neurogenetic explanation observed in this study was due to the addition of information about the neurological basis of the offender's disorder rather than information about its genetic origin.

More recent studies on the effect of genetic information on the perception of moral agency, however, suggest that genetic explanations have no mitigating effect. Using a vignette in which the protagonist committed an impulsive homicide, Appelbaum, Scurich, and Raad (2015) found no evidence that the introduction of genetic evidence influenced judgments about the offender's culpability or appropriate punishment, though it did increase par-

ticipants' fear of the defendant. Results from a related study using vignettes involving less serious offenses exhibited a similar pattern (Scurich & Appelbaum, 2016). These findings are consistent with other research. Cheung and Heine (2015) found no significant difference in sentencing because of genetic explanations of criminal behavior, despite the fact that genetic explanations reduced perceptions of the agent's control over his conduct and increased concerns about his future dangerousness. Fuss, Dressing, and Briken (2015) reached similar conclusions in a study of German judges. They found that genetic evidence did not significantly affect the judges' estimates of appropriate punishment, though it reduced the judges' estimation of the offender's legal responsibility. The genetic evidence, however, did raise the odds that a judge would order involuntary commitment, again suggesting that genetic evidence influences judgments of future dangerousness. Judging from the totality of evidence, it appears that genetic explanations of criminal behavior do not attenuate perceptions of the offender's moral agency. This general pattern of findings is consistent with TDM, on the assumption that explanations of this type do not promote the perception of mentally disordered offenders as victims of harm.

As to whether environmental explanations of behavior attenuate the perception of moral agency, as TDM predicts, the evidence is more mixed. Results from an early study of sentencing in capital cases found that mock jurors were less likely to support a death sentence for a defendant who had been severely abused by his parents as a child (Barnett, Brodsky, & Davis, 2004). These findings were later corroborated by other studies in which the addition of information that an offender had suffered physical or sexual abuse in childhood had a mitigating effect on participants' judgments of appropriate punishment (Barnett, Brodsky, & Price, 2007; Tetterton & Brodsky, 2007). In their review of the literature on juror decision-making in capital cases, covering both mock jury research and research based on posttrial interviews with actual jurors, Sandys, Pruss, and Walsh (2009) concluded that defendants who had experienced traumatic life events outside of their control were less likely to receive a death sentence relative to defendants for whom no mitigating evidence was presented.

Other studies, however, have yielded contrary results. Monterosso, Royzman, and Schwartz (2005) presented participants with vignettes of antisocial conduct accompanied by either a neurological explanation or an explanation involving childhood abuse. They found that ratings of the offender's culpability were significantly higher in the environmental condition. Given the lack of a control condition in their study, it would be wrong to conclude from this finding that the environmental explanation had no mitigating effect. The most that can be safely said is that whatever effect the abuse explanation might have had was less than whatever effect the brain explanation might have had. More recent studies, however, point to a stronger conclusion about the mitigating potential of environmental explanations. For example, Stevenson, Bottoms, and Diamond (2010) examined mock jurors' discussions of evidence that a defendant suffered serious childhood abuse and found that the most frequent kind of juror statement about the abuse urged that it was not mitigating (indeed, some jurors deemed it aggravating). Appelbaum and Scurich (2014) reached similar conclusions with respect to evidence of childhood abuse. In one study, one group of participants received no causal explanation of the offender's impulsivity, while others received

one of three explanations: bad genes, a history of childhood abuse, or a combination of bad genes and childhood abuse. The latter two groups imposed longer sentences relative to the other two groups, despite finding the defendant to be less dangerous. For some reason, participants found the environmental explanation to be aggravating rather than mitigating, contrary to what TDM seems to predict.

Our project continues the investigation into factors relevant to the assignment of blame, punishment, and responsibility in the context of criminal law. The central question is this: In cases where a criminal offender is known to suffer from psychopathy or a similar psychiatric disorder, how are our ordinary judgments of blame, punishment, and responsibility affected by what we know about the etiology of the disorder? Our work is distinctive within the literature on this topic in two respects. First, our vignettes differ in potentially significant ways. Vignettes used in other research on the impact of childhood abuse evidence tend to be short on details, simply stating that the protagonist suffered abuse as a child. Details might matter because of their power to make the causal information salient by engaging participants' imagination. To avoid this potential confound, our vignettes described in some detail the kind of abuse our protagonist suffered.

Second, and perhaps more significantly, whereas most previous studies have investigated ordinary intuitions without a clear theoretical framework, our hypotheses were based on a specific account of moral judgment, namely, the Theory of Dyadic Morality. Our initial hypothesis, derived from this theory, was that ordinary attributions of blame are reduced to a greater extent by evidence of extreme childhood suffering than by evidence of genetic causation. Further, we hypothesized that some environmental explanations mitigate attributions of blame more than others. Specifically, offenders whose disorder is caused by childhood abuse are seen in a less negative light than offenders whose disorder is caused by an unfortunate accident. The underlying idea here is that we view intentional harms as more serious than harms that are similar in all relevant respects apart from being caused accidentally. For example, consider two individuals who have suffered a broken leg, one as the result of an aggravated assault and the other as the result of a skiing accident. Both individuals have been harmed, but the victim of the assault has been doubly harmed, having suffered both a physical injury and the indignity of being disrespected by a wrongdoer. Given the greater moral significance of intentionally caused harm, offenders with a disorder caused by deliberate mistreatment at the hands of others will be perceived more strongly as moral patients and, accordingly, be viewed as less blameworthy relative to offenders whose disorder was caused by accident. Our goal in testing these hypotheses was twofold: first, to explore the effects of etiological information on moral judgment beyond what can be gleaned from the existing literature; and second, to assess the predictive power of TDM as applied to the legal domain.

Study 1: Genes Versus Environment

Our first study used a 3 × 2 factorial design, with the etiology of the agent's brain disorder (genes vs. abuse vs. accident) as the first factor and the type of crime described in the vignette (robbery vs. homicide) as the second factor. The main dependent variables were attributions of blame and punishment. Attributions of free will and "true selfhood"—that is, the extent to which an action

expresses the agent's essential character (Newman, Bloom, & Knobe, 2014)—were also included, as additional indicators of the effects of causal information on the perception of agency.

Our main prediction was the agent would be seen as deserving of more blame and punishment when the etiology of his disorder was genetic rather than environmental. We expected that this same pattern of contrasts would extend to the effects of etiology on the perception of the agent's free will and true selfhood. It was also predicted that, of the two scenarios involving an environmental cause, the agent whose disorder resulted from intentional harm would be viewed less unfavorably than the agent who had been harmed accidentally.

Given the greater moral seriousness of homicide vis-à-vis robbery, it was expected that participants would judge the agent in the homicide vignette more negatively than the agent in the robbery vignette, and that this difference would show up in attributions of blame and punishment. No prediction was made as to whether the difference between types of crime would affect perceptions of free will and true selfhood.

Method

Participants. There were 244 participants (64% female; mean age = 31.3 years) who were recruited from Amazon Mechanical Turk (MTurk). Potential participants were offered $0.20 to complete a brief survey of attitudes toward blame and punishment.[1] Recruitment was limited to individuals living in the United States who had completed at least 50 MTurk tasks with an overall approval rating of 95% or better. Before data collection, the study was certified as exempt by the Institutional Review Board at the University of Missouri.

Materials and procedure. Subjects were randomly assigned to one of two conditions. Depending on condition, subjects read a vignette about either an armed robbery or a homicide.[2] The vignettes read as follows (italics added to indicate contrasting material):

> Brian is 22 years old. He was recently arrested for *armed robbery. Armed with a semiautomatic pistol, Brian entered a bank and ordered a teller to fill a bag with money. Brian grabbed the bag of money and ran out of the bank.* A short time later, Brian was caught by the police.

> Brian is 22 years old. He was recently arrested for *murder. He got into an argument with a store clerk. The argument escalated and Brian assaulted the clerk. Brian repeatedly kicked the clerk in the head after he had fallen to the ground, which caused his death.* A short time later, Brian was caught by the police.

In both conditions, subjects were randomly assigned to one of three continuations of the vignette in which the agent was described as having a brain disorder marked by impairments of empathy and moral judgment.[3] These continuations differed in terms of how the causal history of the disorder was described, namely, whether it resulted from a genetic abnormality, a history of childhood abuse, or an accidental brain injury.

After reading the complete vignette, subjects rated how much blame the agent deserved for the crime (1 = *no blame*, 7 = *a lot of blame*) and how much prison time he deserved for committing it (0–50 years). With respect to the punishment question, participants were instructed to base their decision solely on their estimate of how much blame the offender deserved, regardless of consid-

erations about his future dangerousness and potential to reoffend. They also indicated on a seven-point scale (1 = *strong disagreement*, 7 = *strong agreement*) whether in committing the crime the agent exercised his free will, and on a 5-point scale (1 = *strong disagreement*, 5 = *strong agreement*) whether the crime reflected his true self (characterized as "the deepest, most essential aspects of his personality").[4] (See Appendix for complete materials.)

Results

As predicted, there was a significant effect of causal information on blame, $F(2, 241) = 4.94$, $p = .008$ and punishment, $F(2, 240) = 4.77$, $p = .009$. Pairwise comparisons revealed that the agent was seen as more blameworthy when his brain disorder had a genetic origin than when it resulted from either childhood abuse ($p = .03$, $d = .39$) or an accident ($p = .003$, $d = .47$).[5] Contrary to prediction, however, no significant difference between the two environmental conditions was detected (abuse vs. accident, $p = .41$). With respect to punishment, longer sentences were assigned in the genetic condition relative to the abuse condition ($p = .003$, $d = .49$), but no other significant contrasts were detected (genes vs. accident, $p = .05$; abuse vs. accident, $p = .28$). (See Table 1 for descriptive statistics.)

In addition to the effect of etiological information on blame and punishment ratings, there was an effect of this information on ratings of true selfhood, $F(2, 241) = 3.62$, $p = .03$ and a trend in that direction for free will, $F(2, 236) = 2.92$, $p = .06$. In the case of true selfhood, pairwise comparisons revealed the pattern found with blame and punishment, with ratings of true selfhood higher in the genetic condition relative to both of the environmental conditions (genes vs. abuse, $p = .02$, $d = .36$; genes vs. accident, $p = .02$, $d = .37$), but no significant difference between the environmental conditions (abuse vs. accident, $p = .95$).[6] In the case of free will, pairwise comparisons showed a slightly different pattern, with ratings of free will higher in the genetic condition relative to the accident condition ($p = .02$, $d = .37$), but not the abuse condition ($p = .21$). As elsewhere, the accident and abuse conditions did not differ significantly ($p = .25$).

---

[1] Though MTurk workers as a group are more demographically diverse than college students (Buhrmester, Kwang, & Gosling, 2011), they are not perfectly representative of the population as a whole. MTurk workers in the United States tend to be younger, more highly educated, and more computer literate than average (Ross et al., 2010). As such, they are not perfectly representative of the communities from which jurors are typically drawn—a fact that limits the practical significance of our results.

[2] The homicide vignette was adapted from Monterosso, Royzman, and Schwartz (2005).

[3] Subsequent to data collection, all participant groups were compared on age and gender and found to be equivalent in those respects.

[4] Because of a programming error, different response scales were assigned to the free will and true self questions in Study 1 (7-point and 5-point, respectively). This error was corrected in Study 2.

[5] According to Fisher's Least Significant Difference test, which holds the family wise error rate to $\alpha = .05$ in the special case of three groups, as here (Howell, 2013). All pairwise comparisons reported for Study 1 were analyzed using this test.

[6] Ratings of true selfhood predicted blame ($\beta = .49$, $t(241) = 6.16$, $p < .0001$) and punishment ($\beta = 7.03$, $t(241) = 3.58$, $p = .0004$), consistent with the possibility that the effect of causal information on moral judgment was at least partially mediated by judgment of whether the crime expressed the offender's true self.

Table 1
Mean Judgments in Study 1

Type of judgment	Cause of disorder			Type of crime	
	Genes	Abuse	Accident	Robbery	Homicide
Blame	5.81 (1.28)	5.29 (1.41)	5.10 (1.74)	5.17 (1.52)	5.63 (1.47)
Punishment	23.54 (15.83)	16.04 (14.90)	18.72 (16.53)	10.79 (10.70)	28.02 (15.79)
Free will	6.27 (1.13)	6.00 (1.34)	5.75 (1.61)	5.96 (1.50)	6.06 (1.26)
True self	3.91 (1.08)	3.51 (1.17)	3.50 (1.11)	3.44 (1.20)	3.85 (1.03)

Note. Figures in parentheses are *SD*s.

There was a small but significant effect of crime type on blame, with more blame assigned to the agent in the homicide scenario than in the robbery scenario, $F(1, 242) = 5.80, p = .02, d = .31$. The agent of the homicide was also judged as deserving a much longer prison sentence than the robber, $F(1, 241) = 99.09, p < .001, d = 1.28$. Though the two crimes were seen as equally reflective of the agent's free will, $F(1, 237) = .31, p = .58$, homicide was seen as more expressive of the agent's true self than robbery, $F(1, 242) = 8.12, p = .005, d = .37$. No interaction effects between crime type and cause were observed with any of these variables (all $ps > .4$).

Discussion

The principle of moral typecasting says that the more we see someone as a moral patient (e.g., a victim), the less we see them as a moral agent (e.g., a villain), and conversely (Gray & Wegner, 2009, 2011). Applied to the legal context, the principle suggests that criminal offenders will be judged less negatively when they are perceived as victims of harm relative to offenders who are not so perceived. Accordingly, it predicts that an agent whose criminal behavior is linked to psychopathology will be judged more negatively when the pathology is genetic rather than environmental in origin, because only in the environmental case will the agent be perceived as a victim of harm. This prediction was borne out by the results of Study 1. In two hypothetical crime scenarios, an offender with a brain disorder was perceived as more deserving of blame and punishment when the etiology of his disorder was genetic rather than environmental.

The principle of moral typecasting also yields a secondary prediction, regarding the effects of different types of harm on the perception of moral agency. Because agents of intentional harm tend to be seen as more deserving of blame and punishment (more villain-like) than agents of accidental harm, it follows from the typecasting principle that targets of intentional harm will be seen as more deserving of sympathy and concern (more victim-like) than targets of accidental harm. On this basis we predicted that offenders whose pathology was caused by their having been abused as children would be judged less negatively than offenders whose pathology arose from an accidental brain injury. Contrary to this prediction, offenders who had been abused were judged to be no less deserving of blame and punishment than offenders who had suffered an accident.

In addition to testing these predictions from the Theory of Dyadic Morality, we explored the effects of etiological information on the perception of two other aspects of moral agency: free will and true selfhood. The general pattern observed with respect to these variables echoed the effects of causal information on

blame and punishment, with participants ascribing more free will and true selfhood to the agent with a disorder of genetic origin relative to the agent whose disorder was environmentally caused (though in the case of free will, this contrast was statistically significant only when the cause was accidental). As with attributions of blame and punishment, attributions of free will and true selfhood did not vary as a function of which environmental cause was involved.

With respect to the influence of causal information on participants' judgments, the effect sizes observed in the study were relatively small, ranging from .36 to .49 and averaging .41. Thus, most of the variance seen in participants' judgments of blame and punishment cannot be traced to the effect of etiological information. Though this does not invalidate the conclusions drawn above, it does suggest an important qualification, namely, that the power of causal information to influence how we perceive the moral agency of criminal offenders may be fairly modest.

The results of Study 1 leave open a number of issues. First, they do not settle the question of whether the perception of moral agency is reduced when the agent's disorder is described as having a genetic origin—or even whether such perception is reduced by the introduction of etiological information of any sort. Answering this question requires the addition of a control condition in which the causal history of the agent's disorder is not described at all (a condition that was not included in the design of Study 1). Second, all versions of the vignettes contained some information about the biological basis of the agent's disorder, insofar as the disorder was given a neurological description. No attention was paid to the possible effect of including this information, rather than specifying the disorder in purely psychological terms. Third, the study did not probe perceptions of moral responsibility, an aspect of agency that features prominently in philosophical discussions of free will. To address these limitations, we designed and ran the follow-up study described below.

Study 2: Genes Versus Environment Versus no Etiology

The second study used a 4 × 2 factorial design in which both factors were related to the etiology of a violent crime committed by an agent with a mental disorder. Levels of the first factor corresponded to different causal histories of the disorder (genes vs. abuse vs. accident), plus a control condition in which no causal history was specified. The second factor involved description of the agent's disorder at the mechanistic level, that is, whether or not the description included information about the neurological basis of the disorder. The set of dependent variables from Study

1—judgments of blame, punishment, free will, and true selfhood—was expanded to capture additional dimensions of moral perception, including impressions of the agent's moral responsibility for his crime.

Method

Participants. There were 387 participants (45.5% female; mean age = 33.9 years; 77.1% White, 4.9% Black, 5.4% Hispanic, 9.8% Asian, 2.8% other or nonreporting) who were recruited from Amazon Mechanical Turk. Potential participants were offered $0.25 in exchange for completing a brief survey on attitudes toward blame and punishment. All participants were individuals living in the United States who had completed at least 50 MTurk tasks with an overall approval rating of 95% or better. Before data collection, the study was certified as exempt by the Institutional Review Board at the University of Missouri.

Materials and procedure. All participants read the homicide vignette used in Study 1. Participants were randomly assigned to one of eight groups.[7] Participants in Groups 1–4 read that the agent had a brain disorder involving abnormalities in the prefrontal cortex that impaired his capacity for empathy and moral judgment. In the first three of those groups, respectively, the vignette specified either that the brain disorder was genetic in origin, that it resulted from childhood abuse, or that it resulted from a childhood accident; in the fourth (control) group, there was no information about the etiology of the disorder. Participants in Groups 5–8 read that the agent had a psychological disorder that impaired his capacity for empathy and moral judgment. In the first three of these groups, respectively, participants read that the disorder was because of either a genetic defect, childhood abuse, or a childhood accident; in the fourth (control) group, no etiological information was included.

After reading the vignette, participants judged how much blame the agent deserved for his crime (1 = *no blame at all*, 7 = *a lot of blame*) and how much prison time he should deserve (0–50 years). They also indicated on a 5-point scale (1 = *strong disagreement*, 5 = *strong agreement*) whether the agent was fully morally responsible for the crime, whether he deserved sympathy, and whether there were mitigating factors in the case that reduced his blameworthiness. Using the same scale, participants also indicated whether the agent had acted out of free will and whether the crime reflected his true self. (See Appendix for complete materials.)

Results

As predicted, and consistent with the results of Study 1, there was a significant effect of etiological information on blame, $F(3, 379) = 13.34, p < .001$. Participants assigned more blame to the agent when the cause of his disorder was genetic rather than environmental (genes vs. abuse, $p = .004, d = .49$; genes vs. accident, $p < .001, d = .55$); but the difference between environmental conditions was not significant (abuse vs. accident, $p = .88$).[8] As predicted, the addition of etiological information significantly reduced the attribution of blame in both environmental conditions (abuse vs. control, $p < .001, d = .74$; accident vs. control, $p < .001, d = .77$) but not in the genetic condition (genes vs. control, $p = .64$). (See Table 2 for descriptive statistics.)

There was also a significant effect of etiological information on punishment, $F(3, 379) = 5.14, p = .002$. As in Study 1, the agent

was judged to deserve a longer prison sentence when his disorder was genetically rather than environmentally caused, but this difference reached significance only relative to the accident condition (genes vs. accident, $p = .004, d = .46$; genes vs. abuse, $p = .13$). As before, there was no significant difference between the environmental conditions (abuse vs. accident, $p = .64$). Etiological information was mitigating in the accident condition, but not otherwise (accident vs. control, $p = .01, d = .44$; abuse vs. control, $p = .25$; genes vs. control, $p = .99$).

With respect to other dependent variables in the study, a similar pattern emerged. There was a significant effect of etiological information on judgments of the agent's moral responsibility for the crime, $F(3, 379) = 13.14, p < .001$, the extent to which he deserved sympathy, $F(3, 380) = 15.83, p < .001$, and the presence of mitigating factors in the case, $F(3, 379) = 13.11, p < .001$, as well as judgments of the extent to which the agent exercised his free will, $F(3, 380) = 8.49, p < .001$. The effect of etiological information on ratings of the extent to which the crime expressed the agent's true self did not reach significance, $F(3, 379) = 2.36, p = .07$.

In terms of pairwise comparisons, moral responsibility exhibited the same pattern as was detected in the case of blame. Participants attributed more responsibility to the agent in the genetic condition than in either of the environmental conditions (genes vs. abuse, $p < .001, d = .67$; genes vs. accident, $p < .001, d = .79$), the environmental conditions did not differ significantly (abuse vs. accident, $p = .55$), and etiological information was mitigating in the environmental conditions only (abuse vs. control, $p = .03, d = .42$; accident vs. control, $p < .001, d = .56$; genes vs. control, $p = .50$). The pattern for sympathy was the same but with the direction of effects reversed. Participants expressed less sympathy for the agent in the genetic condition than in either of the environmental conditions (genes vs. abuse, $p < .001, d = .88$; genes vs. accident, $p < .001, d = .52$), the environmental conditions did not differ significantly (abuse vs. accident, $p = .07$), and the addition of etiological information increased sympathy ratings relative to the control in both environmental conditions but not in the genetic condition (abuse vs. control, $p < .001, d = .77$; accident vs. control, $p = .02, d = .42$; genes vs. control, $p = .87$). Ratings of the extent to which the background circumstances of the case were mitigating reflected the same pattern of contrasts observed with ratings of sympathy (genes vs. abuse, $p = .03, d = .39$; genes vs. accident, $p < .001, d = .68$; abuse vs. accident, $p = .11$; abuse vs. control, $p = .006, d = .49$; accident vs. control, $p < .001, d = .78$; genes vs. control, $p = .94$).

In contrast with the results of Study 1, no effect of causal information on ratings of true selfhood was observed, $F(3, 383) = 2.40, p = .07$, but there was an effect of this information on free will, $F(3, 384) = 8.4, p < .001$. Participants attributed more free will in the genetic condition than in either of the environmental conditions, but this contrast reached significance only relative to the accident condition (genes vs. accident, $p < .001, d = .58$; genes vs. abuse, $p = .20$). Free will ratings were slightly higher in the abuse condition than in the accident condition, but the differ-

---

[7] Following data collection, participant groups were compared on all demographic variables (age, gender, and race or ethnicity) and found to be equivalent in those respects.

[8] According to Tukey's test, appropriate for three or more groups, as here (Howell, 2013). All pairwise comparisons reported for Study 2 were analyzed using this test.

Table 2
Mean Judgments in Study 2

Type of judgment	Cause of disorder				Type of disorder	
	Control	Genes	Abuse	Accident	Neurological	Psychological
Blame	6.42 (.90)	6.22 (1.20)	5.62 (1.25)	5.49 (1.45)	5.90 (1.21)	5.98 (1.33)
Punishment	33.30 (15.29)	34.05 (16.91)	29.13 (14.05)	26.47 (15.82)	29.30 (16.07)	32.18 (15.45)
Responsibility	4.26 (.90)	4.44 (.74)	3.87 (.96)	3.69 (1.12)	3.96 (.972)	4.17 (.982)
Sympathy	2.41 (1.13)	2.29 (1.22)	3.26 (1.08)	2.88 (1.12)	2.75 (1.16)	2.66 (1.20)
Mitigation	2.08 (1.15)	2.18 (1.22)	2.63 (1.09)	3.01 (1.23)	2.56 (1.23)	2.39 (1.23)
Free will	4.41 (.72)	4.44 (.75)	4.21 (.74)	3.92 (1.03)	4.16 (.86)	4.33 (.83)
True self	3.81 (.92)	3.91 (1.11)	3.74 (.89)	3.54 (1.04)	3.71 (.98)	3.79 (1.02)

Note. Figures in parentheses are *SD*s.

ence did not reach significance (abuse vs. accident, $p = .06$). The addition of etiological information reduced attributions of free will relative to the control in the accident condition (accident vs. control, $p < .001$, $d = .55$), but not otherwise (abuse vs. control, $p = .33$; genes vs. control, $p = .99$).[9] With respect to ratings of true selfhood, where no effect of etiological information was detected, the difference between the genetic and accident conditions was the only contrast to approach significance (genes vs. accident, $p = .05$; all other $ps > .2$).

Effects of the second, mechanistic factor—that is, whether or not the neural basis of the agent's psychopathology was specified—were limited. Characterizing the disorder in neurological rather than purely psychological terms resulted in a small but significant reduction in ratings of moral responsibility ($F(1, 379) = 5.44$, $p = .020$, $d = .22$) and free will ($F(1, 380) = 4.06$, $p = .045$, $d = .20$). No other effects of this manipulation were detected (blame, $F(1, 379) = .49$, $p = .48$; punishment, $F(1, 379) = 3.34$, $p = .07$; sympathy, $F(1, 380) = .66$, $p = .42$; mitigation, $F(1, 379) = 2.25$, $p = .14$; true self, $F(1, 379) = .74$, $p = .39$).[10]

Discussion

The results of Study 2, like those of Study 1, suggest that the perception of moral agency in psychologically impaired offenders is sensitive to information about the causal history of their pathology. Offenders with a genetically caused impairment are likely to be seen in a more negative light overall than offenders whose impairment has an environmental cause, and offenders whose disorder is characterized as genetic in origin are likely to be judged just as harshly as agents whose disorder is not characterized in etiological terms at all. Agents whose disorder is explained by environmental events, on the other hand, are likely to be judged less harshly than agents with a disorder for which no etiology is offered. These results are broadly consistent with the Theory of Dyadic Morality, assuming that offenders with genetic deficits are not perceived as victims of harm. Harder to square with the theory is the finding, also replicated from Study 1, that offenders who are portrayed as victims of harm are not viewed less unfavorably when the harm in question was inflicted intentionally, rather than by accident. This is surprising, given the greater seriousness of intentional harm relative to accidental harm, and the likely effect of this difference on perceptions of victimhood.

Compared with the results of Study 1, the effects of etiological information on participants' judgments were more pronounced, ranging in size from .39 to .88 (vs. .36 to .49) and averaging .59

(vs. .41). Still, few large effects (i.e., effects of magnitude .8 or greater) were observed, suggesting again that the power of causal explanations to influence how people perceive the agency of criminal offenders should not be overestimated.

Moving away from the Theory of Dyadic Morality, it appears that specifying a neurological basis for an offender's psychopathology has a modest effect on the perception of their moral agency. Though addition of this information did moderate participants' judgments about certain aspects of agency, including moral responsibility and free will, the effect was small in size and not observed across the board. In particular, the manipulation had no effect on attributions of blame or expressions of sympathy.

General Discussion

Results from the two studies presented here show that ordinary judgments of blame, punishment, and other aspects of moral agency are sensitive to information about the etiology of psychological impairments in criminal offenders. In line with the Theory of Dyadic Morality, offenders whose psychopathology was because of environmental causes were seen as less deserving of moral sanction than those whose pathology was genetic in origin; indeed, offenders whose pathology was genetic were judged no less negatively than offenders whose pathology was given no etiological explanation at all. These findings are consistent with prior studies finding no mitigation effect for genetic causal stories on judgments of blame and punishment (Appelbaum, Scurich, & Raad, 2015; Cheung & Heine, 2015; Fuss, Dressing, & Briken, 2015; Scurich & Appelbaum, 2016). However, our results contrast with some previous research on whether evidence of suffering childhood abuse mitigates judgments of blame and punishment (Appelbaum & Scurich, 2014; Stevenson, Bottoms, & Diamond, 2010). We predicted that evidence of childhood abuse would produce greater mitigation in our studies because our vignettes,

[9] Ratings of free will predicted blame ($\beta = .92$, $t(385) = 15.42$, $p < .0001$), punishment ($\beta = 8.01$, $t(385) = 9.10$, $p < .0001$), and responsibility ($\beta = .58$, $t(385) = 11.60$, $p < .0001$), consistent with the possibility that the effect of causal information on moral judgment was at least partially mediated by judgments of whether the offender's actions were freely willed.

[10] There was, however, some evidence of interaction between the mechanistic and causal variables (blame, $F(3, 379) = 2.41$, $p = .07$; sympathy, $F(3, 380) = 4.92$, $p = .002$; mitigation, $F(3, 379) = 2.92$, $p = .03$; all other $ps > .1$).

unlike those used in earlier research, included details about the abuse suffered. This prediction was borne out by our results.

Contrary to hypothesis, however, offenders whose pathology was caused by childhood abuse were seen as no less agentic than those whose pathology resulted from an accident. This finding presents a challenge to the principle of moral typecasting, assuming that offenders with a history of childhood abuse are seen as more victim-like (hence, as less villainous) than offenders with a history of accidental trauma. However, the result may be an experimental artifact, because of the fact that the vignette in the accident condition contained no description of the event that caused the offender's brain injury. Participants may have tacitly assumed that some intentional agency was involved, so that intentional harm was implicated in both cases. This interpretation of the accident scenario could be discouraged by adding details to the vignette that made explicit that the harm suffered by the offender resulted directly from a random event (e.g., a lightning strike). That said, taking intentional agency out of the picture entirely may be impossible, given the human propensity for making sense of natural phenomena in agentic terms, for example, as "acts of God" (Gray & Wegner, 2010).

Concerning our main finding, it might be that the difference in perceived moral agency between the genetic and environmental cases could be given an alternative explanation, in terms of the age of onset of the disorder. In the genetic case, the disorder emerges very early in development, whereas in both of the environmental cases, the disorder does not emerge until childhood. It is conceivable that this difference—a difference of developmental timing (early vs. late), rather than a difference of moral structure (non-harm-involving vs. harm-involving)—is what gave rise to the effect observed. This alternative explanation does not seem especially plausible, but ruling it out would require further empirical work. In particular, the vignettes would need to be revised in such a way that the age of onset of the disorder was invariant across conditions. We predict that the effect of the manipulation would persist despite such invariance, but this prediction has yet to be tested.

Our studies suggest a number of other avenues for future investigation. For one, perhaps the mitigating effects of different causal stories could depend on the type of violence committed by a wrongdoer. Blair, Mitchell, and Blair (2005) distinguish between "reactive" aggression, which is produced in response to a frustrating or threatening event without any further goal, and "instrumental" aggression, which is produced to achieve a specific goal. Perhaps a history of child abuse produces greater mitigating effects when the crime involves reactive aggression rather than instrumental aggression, because participants would be more sympathetic to a victim of childhood abuse who failed to control his temper than one who carefully planned a crime. Based on our results, though, it seems unlikely that the type of aggression matters. The homicide vignette could be interpreted as involving either instrumental or reactive aggression, though the reactive reading seems more natural. The robbery vignette in Study 1, on the other hand, clearly involves instrumental aggression, yet it produced similar results with respect to the different causal stories. Nonetheless, determining whether the mitigating effects of the different causal stories would be affected by the type of violence committed by the perpetrator would require further testing.

Another limitation of our studies is that the causal stories in our vignettes were relatively simple. In each case, the criminal behavior of the protagonist was explained by a diminished capacity for empathy and moral judgment caused by defective genes or childhood trauma.

Reality is more complex. In a capital sentencing trial, the narrative of the defendant's life history might include a wider range of causal factors, including other psychiatric diagnoses, drug or alcohol addiction, and extreme poverty. Our study leaves open whether more complex causal stories involving childhood suffering produce similar mitigating effects on judgments of blame and punishment. For example, participants in a mock jury study of capital sentencing found drug addiction to be an aggravating circumstance (Barnett, Brodsky, & Price, 2007). What about a case in which early childhood abuse causes addiction, and addiction is a significant cause of the criminal behavior? We suspect that the evidence of childhood abuse would still produce a mitigating effect, assuming it was also a cause of the addiction itself. However, further research is required for us to know.

Our findings in Study 2 about the effect of specifying a neural mechanism for an offender's disorder also suggest further avenues for research. Attributing a neural basis to the disorder had only a small mitigating effect on judgments of moral responsibility and no effect on judgments of blame or punishment. This result contrasts to some extent with work by Greene and Cahill (2012), who found that psychiatric diagnostic evidence coupled with evidence from neuropsychological tests and neuroimaging produced greater mitigating effects with respect to punishment than psychiatric diagnostic evidence alone, at least in cases in which the defendant presented a high risk for future dangerousness. Against this background, further investigation into the effect of neuroscientific evidence on judgments of blame and punishment is warranted. A natural extension of our project, for example, would be to add to the design of Study 2 an additional level of the mechanistic factor that included neuroimaging evidence. Given Greene and Cahill's (2012) findings, presenting participants with evidence of this sort might very well influence their judgments of blame and punishment.

As far as the broader significance of our findings is concerned, we note three areas of potential impact: psychology, philosophy, and the law. In closing, we briefly consider each of these areas in turn.

First, it appears that in cases of antisocial behavior associated with psychopathology, the perception of moral agency is sensitive to considerations of the causal history of the pathology. In particular, explaining the pathology as the result of environmental events, such as childhood abuse, tends to reduce judgments of blameworthiness, appropriate punishment, moral responsibility, and the like; explaining the pathology in genetic terms, however, has no such effect. The fact that moral perception appears sensitive to etiological information in this way lends indirect support to the principle of moral typecasting, a central tenet of the Theory of Dyadic Morality (Gray & Wegner, 2009, 2010, 2011). As such, our findings provide further partial confirmation of that theory. They also demonstrate its utility as a tool for illuminating the source of ordinary judgments about blame and punishment in the legal context.

Second, for philosophers engaged in debates about free will and moral responsibility, our results present something of a puzzle for compatibilists and incompatibilists alike. For the incompatibilist, environmental causes should seem to mitigate responsibility, as these are factors beyond the agent's control. However, this effect does not generalize. Genetic causes are not seen as having a mitigating effect, despite the fact that causes of this sort are no more within the agent's control than environmental causes are. Therefore, the incompatibilist's beachhead may not be very robust. Compatibilists have the opposite problem. Whereas incompatibilists need to account for the intuition that genetic explanations do

not diminish responsibility, compatibilists need to account for the intuition that environmental explanations do have a mitigating effect. The challenge facing both incompatibilists and compatibilists, then, is to explain why ordinary judgments about responsibility vary across etiological contexts.

Third, our findings are relevant to the practice of law, particularly in capital cases. It is important not to exaggerate the practical significance of the results reported here, however, especially given the fact that our participants were recruited from MTurk, and MTurk workers as a group are not perfectly representative of the communities from which jurors are drawn (Ross et al., 2010). That said, our research does suggest that consistent with Gray and Wegner's (2011) contention that one way to escape blame is to be a victim, evidence of environmental etiology may be effective for the defense when the etiology of the disorder underlying the defendant's wrongdoing implicates victimhood. Evidence of genetic etiology is a different story. Because genetic behavioral evidence is unlikely to reduce judgments of blame and responsibility, time and resources should be directed toward other strategies, especially if genetic behavioral evidence suggests future dangerousness.

References

Appelbaum, P. S., & Scurich, N. (2014). Impact of behavioral genetic evidence on the adjudication of criminal behavior. *Journal of the American Academy of Psychiatry and the Law, 42*, 91–100.

Appelbaum, P. S., Scurich, N., & Raad, R. (2015). Effects of behavioral genetic evidence on perceptions of criminal responsibility and appropriate punishment. *Psychology, Public Policy, and Law, 21*, 134–144. http://dx.doi.org/10.1037/law0000039

Aspinwall, L. G., Brown, T. R., & Tabery, J. (2012). The double-edged sword: Does biomechanism increase or decrease judges' sentencing of psychopaths? *Science, 337*, 846–849. http://dx.doi.org/10.1126/science.1219569

Barnett, M. E., Brodsky, S. L., & Davis, C. M. (2004). When mitigation evidence makes a difference: Effects of psychological mitigating evidence on sentencing decisions in capital trials. *Behavioral Sciences & the Law, 22*, 751–770. http://dx.doi.org/10.1002/bsl.591

Barnett, M. E., Brodsky, S. L., & Price, J. R. (2007). Differential impact of mitigating evidence in capital case sentencing. *Journal of Forensic Psychology Practice, 7*, 39–45. http://dx.doi.org/10.1300/J158v07n01_04

Blair, R. J. R., Mitchell, D., & Blair, K. (2005). *The psychopath: Emotion and the brain.* Malden, MA: Wiley-Blackwell.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3–5. http://dx.doi.org/10.1177/1745691610393980

Cheung, B. Y., & Heine, S. J. (2015). The double-edged sword of genetic accounts of criminality: Causal attributions from genetic ascriptions affect legal decision making. *Personality and Social Psychology Bulletin, 41*, 1723–1738. http://dx.doi.org/10.1177/0146167215610520

Fuss, J., Dressing, H., & Briken, P. (2015). Neurogenetic evidence in the courtroom: A randomised controlled trial with German judges. *Journal of Medical Genetics, 52*, 730–737. http://dx.doi.org/10.1136/jmedgenet-2015-103284

Gray, K., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology, 96*, 505–520. http://dx.doi.org/10.1037/a0013748

Gray, K., & Wegner, D. M. (2010). Blaming God for our pain: Human suffering and the divine mind. *Personality and Social Psychology Review, 14*, 7–16. http://dx.doi.org/10.1177/1088868309350299

Gray, K., & Wegner, D. M. (2011). To escape blame, don't be a hero—Be a victim. *Journal of Experimental Social Psychology, 47*, 516–519. http://dx.doi.org/10.1016/j.jesp.2010.12.012

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry, 23*, 101–124. http://dx.doi.org/10.1080/1047840X.2012.651387

Greene, E., & Cahill, B. S. (2012). Effects of neuroimaging evidence on mock juror decision making. *Behavioral Sciences & the Law, 30*, 280–296. http://dx.doi.org/10.1002/bsl.1993

Gross, J. (1990, March 31). California execution stayed: State is appealing to high court. *The New York Times.* Retrieved from http://www.nytimes.com

Howell, D. C. (2013). *Fundamental statistics for the behavioral sciences* (8th ed.). Boston, MA: Cengage Learning.

Meyer-Lindenberg, A., Buckholtz, J. W., Kolachana, B. R., Hariri, A., Pezawas, L., Blasi, G., . . . Weinberger, D. R. (2006). Neural mechanisms of genetic risk for impulsivity and violence in humans. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 6269–6274. http://dx.doi.org/10.1073/pnas.0511311103

Monterosso, J., Royzman, E. B., & Schwartz, B. (2005). Explaining away responsibility: Effects of scientific explanation on perceived culpability. *Ethics & Behavior, 15*, 139–158. http://dx.doi.org/10.1207/s15327019eb1502_4

Moore, M. S. (1997). *Placing blame: A general theory of criminal law.* Oxford: New York, NY: Oxford University Press.

Newman, G. E., Bloom, P., & Knobe, J. (2014). Value judgments and the true self. *Personality and Social Psychology Bulletin, 40*, 203–216. http://dx.doi.org/10.1177/0146167213508791

Ross, J., Irani, I., Silberman, M. S., Zaldivar, A., & Tomlinson, B. (2010). Who are the crowdworkers? Shifting demographics in Amazon Mechanical Turk. *CHI EA*, 2863–2872. http://dx.doi.org/10.1145/1753846.1753873

Sandys, M., Pruss, H. C., & Walsh, S. M. (2009). Aggravation and mitigation: Findings and implications. *Journal of Psychiatry & Law, 37*, 189–236. http://dx.doi.org/10.1177/009318530903700204

Schein, C., & Gray, K. (2017). The Theory of Dyadic Morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review.* Advance online publication.

Scurich, N., & Appelbaum, P. (2016). The blunt-edged sword: Genetic explanations of misbehavior neither mitigate nor aggravate punishment. *Journal of Law and the Biosciences, 3*, 140–157. http://dx.doi.org/10.1093/jlb/lsv053

Stevenson, M. C., Bottoms, B. L., & Diamond, S. (2010). Jurors' discussions of a defendant's history of child abuse and alcohol abuse in capital sentencing deliberations. *Psychology, Public Policy, and Law, 16*, 1–38. http://dx.doi.org/10.1037/a0018404

Tetterton, V. S., & Brodsky, S. L. (2007). More is sometimes better: Increased mitigating evidence and sentencing leniency. *Journal of Forensic Psychology Practice, 7*, 79–85. http://dx.doi.org/10.1300/J158v07n03_05

Tiihonen, J., Rautiainen, M. R., Ollila, H. M., Repo-Tiihonen, E., Virkkunen, M., Palotie, A., . . . Paunio, T. (2015). Genetic background of extreme violent behavior. *Molecular Psychiatry, 20*, 786–792. http://dx.doi.org/10.1038/mp.2014.130

Appendix

Vignettes and Probe Questions for Studies 1 and 2

Genetic Condition

Brian does not feel any concern for other people. *Though he was raised in a good home with loving, affectionate parents, Brian was never emotionally attached to them or anyone else, even as a toddler. As he got older, he lied to his family and friends repeatedly, yet he never felt embarrassment when caught in one of his lies.* He developed a cold demeanor and frequently got into fights. Starting when he was a teenager, he committed many crimes, never feeling any guilt or remorse.

Functional brain imaging and other neurological tests of Brian showed abnormalities in the prefrontal cortex and other areas of the brain associated with empathy and moral judgment. These brain abnormalities were probably *genetic in origin.*

Abuse Condition

Brian does not feel any concern for other people. *He was a normal, sweet child who craved affection from his parents. However, Brian suffered severe physical and emotional abuse throughout childhood. His drug-addicted parents whipped him daily with a tree branch. They would lock him in his room for days at a time and then beat him for urinating on the floor. To survive, he ran away from home and learned to fend for himself on the streets.* He developed a cold demeanor and frequently got into fights. Starting when he was a teenager, he committed many crimes, never feeling any guilt or remorse.

Functional brain imaging and other neurological tests of Brian showed abnormalities in the prefrontal cortex and other areas of the brain associated with empathy and moral judgment. These brain abnormalities were probably *caused by the abuse Brian suffered as a child.*

Accident Condition

Brian does not feel any concern for other people. *A freak accident in his childhood left him with a traumatic brain injury. After the accident, he become emotionally detached from everyone around him. As he got older, he lied to his family and friends repeatedly, yet he never felt embarrassment when caught in one of his lies.* He developed a cold demeanor and frequently got into fights. Starting when he was a teenager, he committed many crimes, never feeling any guilt or remorse.

Functional brain imaging and other neurological tests of Brian showed abnormalities in the prefrontal cortex and other areas of the brain associated with empathy and moral judgment. These brain abnormalities were probably *caused by the injury Brian suffered as a result of his accident.*

Control With Neuromechanism (Study 2 Only)

Brian does not feel any concern for other people. Starting when he was a teenager, he committed many crimes, never feeling any guilt or remorse.

Functional brain imaging and other neurological tests of Brian showed abnormalities in the prefrontal cortex and other areas of the brain associated with empathy and moral judgment.

Control Without Neuromechanism (Study 2 Only)

Brian does not feel any concern for other people. Starting when he was a teenager, he committed many crimes, never feeling any guilt or remorse.

Psychological tests of Brian showed impairments of empathy and moral judgment.

Questions for Study 1

- How much blame does Brian deserve for robbing the bank [killing the clerk]? (1 = *no blame at all,* 7 = *a lot of blame*)
- What prison sentence does Brian deserve for robbing the bank [killing the Clerk]? (0–50 years)
- When Brian robbed the bank [killed the clerk], he was acting of his own free will. (1 = *disagree strongly,* 7 = *agree strongly*)
- When Brian robbed the bank [killed the clerk], his behavior was an expression of his true self—it reflected the deepest, most essential aspects of his personality. (1 = *disagree strongly,* 5 = *agree strongly*)

Questions for Study 2

- How much blame does Brian deserve for killing the clerk? (1 = *no blame at all,* 7 = *a lot of blame*)
- How many years in prison does brain deserve for killing the clerk? (0–50 years)
- Brian bears full moral responsibility for his crime. (1 = *strongly disagree,* 5 = *strongly agree*)
- Even though he committed a serious crime, Brian deserves some sympathy. (1 = *strongly disagree,* 5 = *strongly agree*)
- The facts of Brian's case reduce his blameworthiness to some extent. (1 = *strongly disagree,* 5 = *strongly agree*)
- When Brian killed the clerk, he was acting of his own free will. (1 = *disagree strongly,* 5 = *agree strongly*)
- When Brian killed the clerk, his behavior was an expression of his true self—it reflected the deepest, most essential aspects of his personality. (1 = *disagree strongly,* 5 = *agree strongly*)