Psychopathy Increases Perceived Moral Permissibility of Accidents

Liane Young
Boston College

Michael Koenigs, Michael Kruepke, and
Joseph P. Newman
University of Wisconsin–Madison

Psychopaths are notorious for their antisocial and immoral behavior, yet experimental studies have typically failed to identify deficits in their capacities for explicit moral judgment. We tested 20 criminal psychopaths and 25 criminal nonpsychopaths on a moral judgment task featuring hypothetical scenarios that systematically varied an actor's intention and the action's outcome. Participants were instructed to evaluate four classes of actions: accidental harms, attempted harms, intentional harms, and neutral acts. Psychopaths showed a selective difference, compared with nonpsychopaths, in judging accidents, where one person harmed another unintentionally. Specifically, psychopaths judged these actions to be more morally permissible. We suggest that this pattern reflects psychopaths' failure to appreciate the emotional aspect of the victim's experience of harm. These findings provide direct evidence of abnormal moral judgment in psychopathy.

Keywords: psychopathy, moral judgment, accident, intention, emotion

Supplemental materials: http://dx.doi.org/10.1037/a0027489.supp

Psychopathy is associated with immoral and criminal behavior (Hare, 1993). Such gross deviance in moral and social conduct might suggest underlying abnormalities in moral competence—the basic ability to distinguish right from wrong (Mikhail, 2007). Indeed, psychopathic traits in adults and adolescents have been linked to a deficit in making the "moral-conventional distinction," that is, the ability to distinguish violations of others' welfare (e.g., hitting a classmate) from violations of social convention (e.g., wearing pajamas to class) (Blair, 1995, 1997). Research on psychopathy has also revealed deficits in moral reasoning (Fodor, 1973; Jurkovic & Prentice, 1977) and conditional reasoning related to social contracts and precautionary rules (Ermer & Kiehl, 2010).

Recent evidence points to possible psychological and neurobiological sources of abnormal moral and social function, including impairments in social-emotional processing. For example, psychopaths lack normal implicit associations between unpleasant words

and violent words (N. S. Gray, MacCulloch, Smith, Morris, & Snowden, 2003). Certain psychopathic traits have also been associated with hyper-reactivity of the dopaminergic reward system in response to pharmacological and monetary reinforcers (Buckholtz et al., 2010), hypo-reactivity of the amygdala during emotional moral decision-making (Glenn, Raine, & Schug, 2009), atypical activity in the amygdala and ventromedial prefrontal cortex during the viewing of morally salient scenes (Harenski, Harenski, Shane, & Kiehl, 2010), and reduced structural and functional connectivity between amygdala and ventromedial prefrontal cortex (Motzkin, Newman, Kiehl, & Koenigs, 2011).

These landmark studies investigate cognitive and emotional processes that are likely to be important for and related to moral judgment (Bloom, 2011; Schaich-Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Moll, de Oliveira-Souza, Bramati, & Grafman, 2002; Nichols, 2002; Valdesolo & DeSteno, 2006; Wheatley & Haidt, 2005). However, they do not focus on explicit moral judgment directly; for example, they do not test psychopaths' capacity to judge agents' actions as morally right or wrong, or the agents themselves as morally good or bad. Notably, recent studies that have directly investigated the capacity for explicit moral judgment have failed to demonstrate consistent deficits in psychopaths, including recent studies probing psychopaths' responses to moral dilemmas (e.g., the permissibility of harming one to save many) (Cima, Tonnaer, & Hauser, 2010; Glenn, Raine, Schug, Young, & Hauser, 2009; Koenigs, Kruepke, Zeier, & Newman, in press). A number of early studies relying on tasks measuring moral justification (Kohlberg, 1969) also discovered no deficits in psychopathy (Lee & Prentice, 1988; Link, Scherer, & Byrne, 1977; O'Kane, Fawcett, & Blackburn, 1996; Simon, Holzberg, & Unger, 1951; Trevethan & Walker, 1989). Moreover, according to a recent study on incarcerated offenders, even the classic finding that psychopaths are unable to make the moral-

conventional distinction may be subject to revision; no relationship was found between total psychopathy score and task performance (Aharoni, Sinnott-Armstrong, & Kiehl, in press).

These mixed findings have been taken to reflect a puzzling dissociation between normal moral knowledge, on the one hand, and abnormally immoral behavior, on the other (Cima et al., 2010; Cleckley, 1941). In other words, psychopaths appear competent to make the same moral distinctions as everyone else in their explicit moral judgments, in spite of their actual behavior. On this proposal, gross abnormalities in the moral behavior of psychopaths do not stem from impairments in underlying moral knowledge (Cima et al., 2010). By contrast, we suggest that the psychological deficits that have been explored in prior research on psychopathy (e.g., insensitivity to the suffering of others) lead not only to abnormal moral motivation and behavior but also to measurable differences in explicit moral judgments that rely on the same processes.

Here we provide novel evidence for abnormal moral judgments in psychopathy. Rather than relying on prior paradigms that featured only *intentional* harms, we included *accidental* harms as well; this approach allows us to distinguish judgments based on harmful intentions from judgments based on harmful outcomes. We presented participants with scenarios depicting harmful or neutral outcomes, caused by agents with harmful or neutral intentions. More specifically, participants read and responded to four types of 48 moral scenarios (cf. Young, Bechara, et al., 2010), which systematically varied the actor's intention and the action's outcome: (1) accidental harms (neutral intention, harmful outcome), (2) attempted harms (harmful intention, neutral outcome), (3) intentional harms (harmful intention, harmful outcome), and (4) neutral acts (neutral intention, neutral outcome). For example, in the case of an accident, a person might unintentionally kill her neighbor's child by offering him peanuts without knowing about his fatal peanut allergy.

In the case of accidental harms, normal moral and legal judgments usually converge on assigning some blame to the responsible agent, in spite of the agent's innocent intentions, on the grounds that the agent actually caused, in some cases, serious harm (e.g., manslaughter) (Cushman, 2008; Cushman, Dreber, Wang, & Costa, 2009; Killen, Lynn Mulvey, Richardson, Jampol, & Woodward, 2011; Young, Cushman, Hauser, & Saxe, 2007; Young, Nichols, & Saxe, 2010). Even when full forgiveness is theoretically warranted in the absence of negligence or recklessness (e.g., there was absolutely no way that the agent, or anyone in the agent's position, could have anticipated any harm to the victim), people still tend to judge the action as somewhat morally wrong, simply in response to the outcome (Young, Nichols, et al., 2010). This intuitive response to accidents reflects cognitive conflict of

the sort that characterizes much of moral psychology and especially moral dilemmas (Cushman & Young, 2009; Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene & Paxton, 2009). In particular, evaluating accidents elicits competing psychological responses—one based on the emotionally salient harmful outcome and the other based on a representation of the agent's neutral mental state, or innocent intention (Cushman, 2008; Young et al., 2007). It is precisely this cognitive conflict that we hypothesized to be attenuated in psychopathy, as a direct result of psychopaths' reduced sensitivity to the pain and suffering of others (Aharoni, Antonenko, & Kiehl, 2011; Blair, 1995; Glenn, Iyer, Graham, Koleva, & Haidt, 2009; K. Gray, Jenkins, Heberlein, & Wegner, 2011; N. S. Gray et al., 2003).

We tested 20 criminal psychopaths and 25 criminal but nonpsychopathic comparison participants (Table 1). We hypothesized that psychopaths' moral judgments would reflect abnormal (reduced) sensitivity to harmful outcomes and that therefore, in the case of accidents, their moral judgments would be based primarily on information about the agent's neutral intent. In sum, we predicted that psychopaths would deliver more lenient moral judgments specifically for accidents.

Method

Participants

Participants were Caucasian male inmates recruited from a medium-security Wisconsin correctional institution. Inmates were eligible if they met the following criteria: under 45 years of age, no history of psychosis or bipolar disorder, and not currently taking psychotropic medications. A total of 64 inmates met the inclusion criteria and participated in all study procedures. Informed consent was obtained both orally and in writing.

The Psychopathy Checklist-Revised (PCL-R) (Hare, 2003) was used to assess psychopathy. The PCL-R assessment involves a 60–90 minute interview and file review to obtain information used to rate 20 psychopathy-related items as 0, 1, or 2, depending on the degree to which each trait characterizes the individual. A substantial literature supports the reliability and validity of PCL-R assessments with incarcerated offenders (Hare, 2003). To evaluate interrater reliability, a second rater who was present during interviews provided independent PCL-R ratings for eight inmates. The intraclass correlation coefficient was 0.85. The sample was restricted to Caucasian inmates in light of previous studies demonstrating that cognitive and affective performance deficits observed among Caucasian psychopathic inmates (based on the

Table 1
Participant Group Characteristics

	PCL-R total	PCL-R F1	PCL-R F2	Age	Est. IQ
Psychopaths ($n = 20$)	31.7 (1.6)	11.4 (1.7)	17.0 (1.5)	32.5 (6.0)	101.6 (9.8)
Intermediate ($n = 19$)	25.3 (2.5)	9.1 (1.7)	14.0 (2.4)	28.5 (6.3)	102.8 (10.0)
Non-psychopaths ($n = 25$)	14.2 (3.4)	5.0 (3.0)	7.6 (2.7)	32.0 (7.1)	102.1 (11.8)

Note. Est. IQ = Estimated IQ based on the Shipley Institute of Living Scale (Zachary, 1986). For each group, means are presented with standard deviations in parentheses. The three participant groups did not significantly differ with respect to age ($F = 2.2$, $p = .13$) or estimated IQ ($F = 0.06$, $p = .94$).

PCL-R) did not replicate among African-American psychopathic inmates (Baskin-Sommers, Vitale, MacCoon, & Newman, in press; Kosson, Smith, & Newman, 1990; Lorenz & Newman, 2002).

Participants were classified as psychopathic if their PCL-R scores were 30 or greater ($n = 20$) and nonpsychopathic if their PCL-R scores were 20 or less ($n = 25$) (Hare, 2003). Psychopaths ($M = 32.5\ SD = 6.0$) and nonpsychopaths ($M = 32.0\ SD = 7.1$) did not differ in age, $t(43) = -.25$, $p = 0.80$, or IQ (psychopath $M = 101.6\ SD = 9.8$; nonpsychopath $M = 102.1\ SD = 11.8$; $t(42) = 0.16$, $p = 0.88$) (Table 1). A secondary analysis examined participants in an intermediate psychopathy group, with PCL-R scores of 21–29 ($n = 19$).

Procedure

After completing the PCL-R and IQ assessments, participants then completed the moral judgment task (i.e., 48 hypothetical moral scenarios), a previously published task used in a number of prior studies on both neurotypical and patient populations (Miller et al., 2010; Moran et al., 2011; Young, Bechara, et al., 2010; Young, Camprodon, Hauser, Pascual-Leone, & Saxe, 2010; Young & Saxe, 2008, 2009a). In some cases, the original scenario language was modified to be more easily understood by inmates with limited reading skills. The moral scenarios described harmful or neutral outcomes and intentions: accidental harms (neutral intention, harmful outcome), attempted harms (harmful intention, neutral outcome), intentional harms (harmful intention, harmful outcome), and neutral acts (neutral intention, neutral outcome). Thus, each of the 48 scenarios appeared in all four conditions, across participants, although each participant saw only one version (condition) of each scenario (available online as supplemental material).

In one example of an *accidental harm*, participants read the following: (1) Grace and her friend are taking a tour of a chemical plant. When Grace goes over to the coffee machine to pour some coffee, Grace's friend asks for some sugar in hers. There is white powder in a container by the coffee. (2) The white powder is a very toxic substance left behind by a scientist, and therefore deadly when ingested in any form (*information foreshadowing negative outcome*). (3) The container is labeled "sugar," so Grace believes that the white powder by the coffee is sugar left out by the kitchen staff (*neutral intent*). (4) Grace puts the substance in her friend's coffee. Her friend drinks the coffee and dies (*negative outcome*).

Other participants read the *attempted harm* version of this same "coffee" scenario, which began with the same background information, followed by these facts: (2) The white powder is just the regular sugar that the scientists use every day, and is therefore perfectly safe to have in coffee (*information foreshadowing neutral outcome*). (3) The container is labeled "toxic," so Grace believes that the white powder is toxic substance left behind by a scientist (*negative intent*). (4) Grace puts the substance in her friend's coffee. Her friend drinks the coffee and is fine (*neutral outcome*).

Half of the scenarios presented information foreshadowing the outcome first, the other half information about the intention first. Order did not interact with any term involving group ($p > .34$). In addition, we presented four different scenario-condition pairings, assigned randomly across participants. Pairing also did not interact

with any term involving group ($p > .05$). For example, one participant might see the accidental harm version of the "coffee" and the attempted harm version of the "latex" scenario, whereas another participant might see the accidental harm version of the "latex" scenario and the attempted harm version of the "coffee" scenario. Thus, different participants saw different items (scenarios) in different conditions (accidental, attempted, intentional, neutral). However, we were able to determine the internal consistency within each condition across scenario-condition pairings (standardized Cronbach's alpha for intentional harm: 0.75; attempted harm: 0.84; accidental harm: 0.85; neutral act: 0.65). In addition, as we describe below in item-wise analyses, the effects do not appear to be driven by any particular item (see also Figure 1 and Figure 2).

Each scenario was presented on a single sheet of paper, and participants circled a number (1–7) to rate the moral permissibility of the action in the scenario (1 = morally forbidden, 7 = morally acceptable). There was no time limit for reading the scenario description or responding to the question.

Results

Subject-Wise Analyses

A 2 (intent: harmful vs. neutral) × 2 (outcome: harmful vs. neutral) × 2 (group: psychopath vs. nonpsychopath) mixed
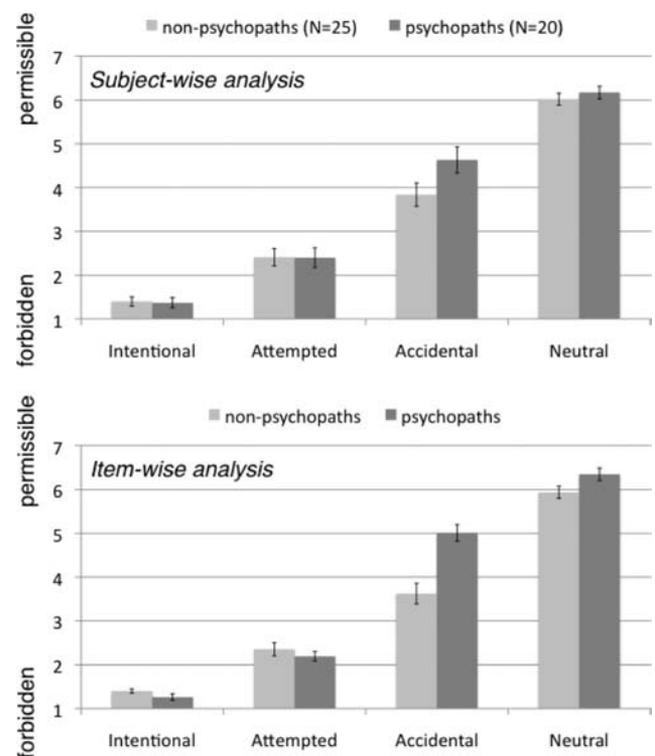

Figure 1. Moral judgments on a seven-point scale (1 = *morally forbidden*; 7 = *morally permissible*), analyzed across subjects (top) and across items (bottom). Psychopaths (dark bars) judge accidental harms as more morally permissible than non-psychopaths (light bars). Error bars indicate standard error of the differences.
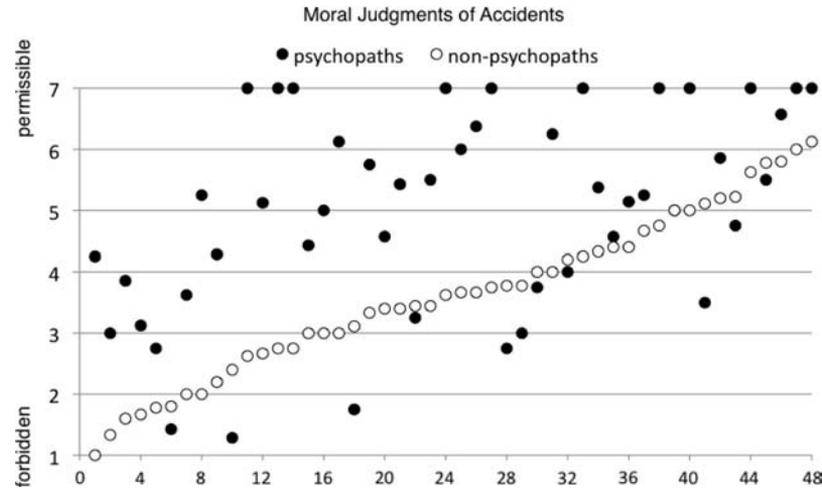
Figure 2. Moral judgments of accidents on a seven-point scale (1 = *morally forbidden*; 7 = *morally permissible*). On the *x* axis, scenarios are in order of increasing moral permissibility as judged by non-psychopaths. Psychopaths (filled-in circles) judge accidental harms more permissible than non-psychopaths (hollow circles) on 37 of 48 scenarios.

effects ANOVA of participants' moral judgments yielded main effects of intent and outcome. As expected, across both groups, actions performed with harmful intentions were judged to be less morally permissible than actions performed with neutral intentions, $F(1, 43) = 351$, $p < .001$, partial $\eta^2 = 0.89$. Also, actions resulting in harmful outcomes were judged to be less morally permissible than actions resulting in neutral outcomes, $F(1, 43) = 121$, $p < .001$, partial $\eta^2 = 0.74$ (see Table 2 for descriptive statistics).

As the key test of our hypothesis, this analysis revealed the critical three-way intent by outcome by group interaction, $F(1,$ 43) = 5.8, $p = .02$, partial $\eta^2 = 0.12$ (Figure 1), revealing selective differences between participant groups depending on the specific class of moral action. No other main effect or interaction was significant ($p > 0.16$). Specifically, as predicted, psychopaths showed a selective abnormality in evaluating accidental harms; psychopaths judged accidents—harmful outcomes brought about by agents with neutral intentions—to be more morally permissible, compared with nonpsychopaths, $t(43) = -1.9$, $p = 0.05$. Notably, psychopaths and nonpsychopath participants showed no differences on any of the other conditions: attempted harms, intentional harms, or neutral acts ($p > 0.46$).

Table 2
Descriptive Statistics for Subject-Wise and Item-Wise Analyses

Condition	Range	Min, Max	Mean (*SE*)	*SD*
Subject-wise				
Non-psychopaths (*n* = 25)				
Intentional	2	1, 3	1.40 (.086)	0.431
Attempted	4	1, 5	2.41 (.192)	0.961
Accidental	5	1, 6	3.84 (.269)	1.35
Neutral	3	1, 7	6.02 (.136)	0.68
Psychopaths (*n* = 20)				
Intentional	2	1, 3	1.37 (.137)	0.611
Attempted	4	1, 5	2.40 (.235)	1.049
Accidental	5	2, 7	4.63 (.293)	1.312
Neutral	2	5, 7	6.17 (.144)	0.643
Item-wise (*n* = 48 scenarios)				
Non-psychopaths				
Intentional	2	1.00, 3.00	1.40 (.075)	0.519
Attempted	3.55	1.25, 4.80	2.35 (.107)	0.742
Accidental	5.13	1.00, 6.13	3.62 (.191)	1.322
Neutral	4	3.00, 7.00	5.94 (.143)	0.989
Psychopaths				
Intentional	1.25	1.00, 2.25	1.26 (.050)	0.356
Attempted	4	1.00, 5.00	2.19 (.152)	1.052
Accidental	5.71	1.29, 7.00	5.01 (.236)	1.638
Neutral	4.25	2.75, 7.00	6.35 (.138)	0.961

Item-Wise Analysis

To determine whether the effects above were attributable to any particular items (i.e., scenarios), we conducted an item-wise analysis, treating each of the 48 scenarios as a case (as opposed to each participant as a case). A 2 (intent: harmful vs. neutral) $\times$ 2 (outcome: harmful vs. neutral) $\times$ 2 (group: psychopath vs. nonpsychopath) repeated measures ANOVA yielded main effects of intent and outcome. Again, actions performed with harmful intentions were judged to be less morally permissible than actions performed with neutral intentions, $F(1, 47) = 796.2$, $p < .001$, partial $\eta^2 = 0.94$. Also, actions resulting in harmful outcomes were judged to be less morally permissible than actions resulting in neutral outcomes, $F(1, 47) = 216.6$, $p < .001$, partial $\eta^2 = 0.82$ (see Table 2 for descriptive statistics).

Importantly, the item-wise analysis replicated the key three-way intent by outcome by group interaction, $F(1, 47) = 8.2$, $p = .006$, partial $\eta^2 = 0.15$ (Figure 1). Psychopaths judged accidents to be more morally permissible, compared with nonpsychopaths, $t(47) = -6.27$, $p < 0.001$ (Figure 2). This analysis therefore shows that the key results emerge not only across participants but across items in the moral judgment task.

Intermediate Psychopathy Group

The permissibility ratings delivered by intermediate psychopathy individuals closely matched the nonpsychopaths for each scenario type (all p values > 0.54 in paired-samples t tests; Figure S2 and Figure S3, available online as supplemental material). Compared with psychopaths, the intermediate group made marginally lower permissibility ratings for accidents, $t = -1.9$, $p = .06$, but delivered similar ratings for all other scenario types (all p values > 0.35). Thus, the increased moral permissibility of accidental harms appears to be specific to the high psychopathy group (Figure S2 and Figure S3, available online as supplemental material).

Factor and Facet Scores

The inclusion of intermediate psychopathy scores affords the opportunity to examine correlations between psychopathy scores and moral judgments of accidental harms, across the entire range of psychopathy severity. As expected from the group data, PCL-R total score and moral judgments of accidental harms showed a weak positive correlation ($r = .22$, $p = .08$). Similar correlations were obtained for both Factor 1 score ($r = .19$, $p = .14$) and Factor 2 score ($r = .19$, $p = .14$). Factor 1 and Factor 2 scores were significantly correlated with each other in this participant sample ($r = .53$, $p < 0.001$). Partial correlations indicate that neither Factor 1 nor Factor 2 scores were uniquely associated with moral judgments of accidents (Factor 1: $r_{pc} = 0.10$, $p = .42$; Factor 2: $r_{pc} = 0.11$, $p = .41$). Thus, the relationship between psychopathy and moral judgments appears to reflect the common (i.e., overlapping) variance between PCL-R Factors 1 and 2.

The four-facet model of psychopathy (Hare, 2003) yielded more selective results. We report here the correlations with moral judgments of accidents: Facet 1 ($r = .24$, $p = .06$), Facet 2 ($r = .08$, $p = .54$), Facet 3 ($r = .12$, $p = .37$), Facet 4 ($r = .20$, $p = .11$). These data indicate that the strongest relationship emerged between moral judgments of accidents and Facet 1 items of the PCL-R (e.g., glibness, grandiosity, lying, and manipulativeness). Again, however, examination of partial correlations revealed that none of the unique effects approached statistical significance (all $r_{pc} < 0.17$; all $p > .21$), indicating that our findings relate most closely to the general psychopathy construct rather than any of its individual facets.

Consistency of Results

It is important to note that the key difference between psychopaths and nonpsychopaths in moral judgments of accidents is not attributable to the presence of one or two outliers, either outlying items (e.g., scenarios) or outlying participants. First, a substantial majority of items (37 of 48) elicited the predicted effect—psychopaths judged accidents as more morally permissible than nonpsychopaths (Figure 2). Second, a substantial majority of psychopathic participants showed the predicted effect; 15 of 20 psychopaths judged accidents to be more permissible, compared with the mean moral judgment of accidents of nonpsychopaths (Figure S3). Thus, we observed the same robust effects in both the subject-wise and the item-wise analyses reported above.

Discussion

The current findings provide novel evidence for abnormal moral cognition in psychopathy at the level of explicit judgments. This difference emerged specifically for moral judgments of accidents, which in the present study consisted of one person unknowingly causing harm (i.e., severe injury or even death) to another person. Although accidents, by definition, occur in spite of innocent intentions, people normally assign some amount of blame to the agent on account of the actual damage done (Cushman, 2008). As predicted, psychopaths in the present study judged accidental harms to be more morally permissible (less blameworthy) than nonpsychopaths but showed no such difference for any other moral condition (i.e., intentional harms, attempted harms, neutral acts). We discuss possible reasons for this selective decision-making deficit below.

Harmful Intent Versus Accidents

Notably, prior studies on explicit moral judgment that demonstrated no consistent behavioral differences between psychopaths and nonpsychopaths focused on *intentional* harms, in the context of moral dilemmas (Cima et al., 2010; Glenn, Raine, Schug, et al., 2009). Meanwhile, studies that did reveal a difference between psychopaths and nonpsychopaths investigated not the capacity to distinguish moral right from wrong but the capacity to make a distinction between moral and putatively nonmoral acts, for example, to differentiate moral violations from social-conventional violations (Blair, 1995). Furthermore, these tasks featured only *intentional* violations—in both the moral domain and the social-conventional domain. By contrast, in the present paradigm, the critical difference between psychopaths and nonpsychopaths emerged uniquely for moral judgment of *unintended* acts (e.g., harms).

Indeed, prior studies have shown that theory of mind, including the ability to reason about an agent's intent, is largely preserved in psychopaths (Blair et al., 1996; Dolan & Fullam, 1999; Richell et

al., 2003). What may be lacking instead is a normal emotional response to a victim's experience of harm, in our case, at the hands of an innocent other (Glenn, Raine, & Schug, 2009; K. Gray et al., 2011; Harenski et al., 2010). While emotional responding may be critical for all moral judgments, psychopaths may be able to deliver apparently normal judgments of intentional harms and even attempted harms based on an intact understanding of the malicious intent of the actor. This strategy, however, fails to produce normal moral judgments in the case of accidents, where the judgment likely depends (at least partially) on an emotional response to the victim's unintended pain. Notably, then, a moral judgment task that probes the capacity to evaluate only intentional harms might not uncover any deficits even in criminal offenders with relatively high levels of psychopathy (Cima et al., 2010; Glenn, Raine, Schug, et al., 2009).

Importantly, the behavioral pattern observed for psychopaths was highly specific to this unique population—the same moral judgment task administered to individuals with autism (Moran et al., 2011), complete and partial callosotomy patients (Miller et al., 2010), and patients with lesions to ventromedial prefrontal cortex (VMPC) (Young, Bechara, et al., 2010) yielded distinct moral judgment abnormalities. For instance, VMPC patients did not judge accidents any differently but instead judged attempted harms, including diverse forms of attempted murder, as more morally permissible, possibly because of a defect in processing abstract mental state information. By contrast, psychopaths in the present study judged accidental harms (e.g., manslaughter) as more morally permissible, delivering the more intent-based moral judgment. This unique pattern may be attributable to psychopaths' specific failure to intuitively appreciate the emotional aspect of a victim's experience (Bartels & Pizarro, in press; K. Gray et al., 2011).

Are Psychopaths More Rational?

In a sense, psychopathy in the current study was associated with the more rational response—greater forgiveness for an accident. Theory of mind is thought to provide critical input to moral judgment (Cushman, 2008; Knobe, 2005; Mikhail, 2007). Indeed, before young children fully develop a mature theory of mind—presumably an essential capacity for inhibiting a prepotent response to an emotionally salient but unintended bad outcome (Young & Saxe, 2009b)—young children are much harsher on accidents than adults (Baird & Astington, 2004; Killen et al., 2011). That is, young children's moral judgments are disproportionately based on their responses to the harmful outcome alone. A similar pattern has been found in high-functioning autism (Moran et al., 2011); adults with autism, who show impairments in theory of mind, assign more moral blame to accidents, indicating a greater weighting of the salient negative outcome. Importantly, even neurotypical adults (like the comparison participants of the current study) assign some blame and even punishment for accidents, which is taken to reflect an intuitive rather than reasoned response to the salient outcome (Cushman et al., 2009). Psychopaths, however, appear to be on the opposite end of this moral spectrum, assigning relatively *little blame* to accidents, with healthy adults in the middle of the spectrum (e.g., assigning *some blame* to accidents), and children and individuals with autism on the other end (e.g., assigning *substantial blame*).

Prior research has revealed individual differences among healthy adults in moral judgments of accidents (Cohen & Rozin, 2001; Sargent, 2004; Young & Saxe, 2009b). In one study, participants who showed greater recruitment of brain regions implicated in theory of mind or mental state reasoning (e.g., right temporo-parietal junction, RTPJ; Jenkins & Mitchell, 2010; Perner, Aichhorn, Kronbichler, Staffen, & Ladurner, 2006; Saxe & Kanwisher, 2003; Young, Camprodon, et al., 2010) were more likely to forgive accidents in their behavioral judgments, plausibly reflecting greater consideration of the agent's innocent intention (vs. the harmful outcome), compared with participants who showed lower RTPJ activation (Young & Saxe, 2009b). The present findings suggest multiple routes to the same behavioral pattern, that is, greater forgiveness of accidents. On the one hand, enhanced theory of mind and consideration of the agent's innocent intention may lead to forgiving accidents; on the other hand, reduced sensitivity to the victim's experience of pain and suffering, as we have hypothesized to be the case in the current psychopathic participants, could also lead to the same behavioral response.

Caveats Concerning Psychopathy

We consider two important caveats before discussing possible interpretations of the psychopathic participants' abnormal performance in the next section. First, the current moral judgment task was not designed to test specific psychological models of psychopathy (i.e., the task does not include sufficient manipulations to do so).

Second, our primary findings were obtained for categorically assessed psychopathy; that is, participants were classified as psychopathic if their PCL-R scores were 30 or greater and nonpsychopathic if their PCL-R scores were 20 or less (Hare, 2003). Moral judgments delivered by intermediate psychopathy individuals closely matched the nonpsychopaths across each type of moral scenario. We do note, however, diverse pathways lead to antisocial behavior that may result in elevated PCL-R scores (Skeem, Johansson, Andershed, Kerr, & Eno Louden, 2007). For example, intermediate scores may reflect distinct processes (e.g., executive funtion) that are less related to interpersonal callousness (Morgan & Lilienfeld, 2000). In addition, when we examined correlations between psychopathy scores (PCL-R factors and facets) and moral judgments of accidental harms, across the entire range of psychopathy severity, we did not find any significant effects (with the exception of a marginal finding for Facet 1 of the PCL-R, e.g., glibness, grandiosity, lying, and manipulativeness). Additional research using a larger sample will be necessary to determine whether abnormal judgments of accidental harm is reliably specific to categorically defined psychopathy.

Alternative Interpretations

While the behavioral pattern observed in the current psychopathic participants may be primarily attributable to reduced emotional responding to the salient harmful outcome, an alternative explanation appeals additionally to differential attentional processing. A wealth of evidence indicates that decision-making in psychopathic offenders is relatively unaffected by contextual cues that are incongruent with their response inclinations (Hiatt, Schmitt, &

Newman, 2004; Newman & Baskin-Sommers, 2011). Because accidental harms typically feature conflicting information (i.e., actions and outcomes that are incongruent with stated intentions), psychopathic offenders may have focused primarily on agents' intentions and neglected outcomes.

On this attention-based account, psychopaths deliver more lenient judgments of accidents because they are less likely to reconcile conflicting (harmful) outcome information and more likely to take the stated (innocent) intentions at face value. Prior research suggests that psychologically "normal" participants often infer intentions that are consistent with agents' actions, even when agents are explicitly described as holding different intentions (Wertz & German, 2007). In one example, (1) a story protagonist, Mary, is described as putting her hairdryer next to her perfume in the drawer and then exiting the room; (2) Gina moves the hairdryer to the cabinet; and (3) Mary comes back into the room *for her hairdryer* and goes directly to the drawer. When asked why Mary went to the drawer, adult participants commonly explained: "Because she wanted to get her perfume" (even though Mary was stated as returning *for her hairdryer*). These results suggest that "actions speak louder than words," including words about internal, unobservable intentions (Wertz & German, 2007).

This normal tendency to align agents' intentions with their actions and outcomes may be especially pronounced in specific moral contexts. For example, participants exhibit hindsight bias when judging agents who cause *bad* outcomes (Royzman & Kumar, 2004). Indeed, for a subset of the scenarios featured in the current study, neurotypical adults were more likely to judge that agents were not justified in holding false beliefs when these false beliefs led to bad (but not good) outcomes, for example, Grace *shown have known* that the powder was in fact poison (when she falsely believed that it was sugar) (Young, Nichols, et al., 2010). In other words, information about bad, harmful outcomes led participants to second-guess agents' innocent intentions (e.g., were they really that innocent?). According to convergent findings, participants often infer that agents who cause (unintended) bad outcomes hold less-than-noble intentions (Alicke, 2000; Knobe, 2005). Psychopaths, however, may be less likely to "integrate" intentions and outcomes in this fashion.

Based on these prior findings, we suggest it is the increased affective salience of bad outcomes or actions that make it especially difficult for psychologically "normal" participants to accept the stated (innocent) intention. Crucially, this conflict is absent for failed attempts to harm where intent (bad) and outcome (neutral) are also in conflict; that is, participants are unlikely to infer that these agents have neutral intentions simply because a neutral outcome occurred. Notably, psychopaths in the current study diverged from nonpsychopaths only in the case of accidents—not attempts. Regardless of whether psychopaths' apparent insensitivity to the harmful outcomes in the accidental harm scenarios, specifically, is primarily attributable to an attention-based deficit or to an emotion-based deficit or both, the findings here demonstrate that psychopathy is indeed associated with measurable differences in moral judgment. Future work is required to disentangle these possibilities, but, based on the current and prior findings, we suggest that both kinds of deficits may be at play. Owing to psychopaths' cognitive-affective deficits, harmful outcomes may become less salient; consequently, neutral intentions may become more primary, further reducing the impact of harm.

Conclusions

The current findings may bear on important debates within philosophy and jurisprudence. Legal scholars have asked, for example, whether psychopaths can be held legally responsible for misconduct if they lack moral knowledge (Aharoni, Funk, Sinnott-Armstrong, & Gazzaniga, 2008; Morse, 2008). A separate philosophical question is whether a psychological deficit actually leads psychopaths to make normatively superior judgments (Bartels & Pizarro, in press), to forgive accidents where others ought to but cannot because of their intuitive response to the harm (Prinz, 2004; Prinz, 2008). Indeed, normal moral judgments may be colored by people's automatic emotional responses to harms, both intended and unintended. Yet, we may crucially rely on these gut responses to regulate our own moral behavior and to inhibit ourselves from doing harm to others (Blair, 1995).

References

Aharoni, E., Antonenko, O., & Kiehl, K. A. (2011). Disparities in the moral intuitions of criminal offenders: The role of psychopathy. *Journal of Research in Personality, 45,* 322–327. doi:10.1016/j.jrp.2011.02.005

Aharoni, E., Funk, C., Sinnott-Armstrong, W., & Gazzaniga, M. (2008). Can neurological evidence help courts assess criminal responsibility? Lessons from law and neuroscience. *Annals of the New York Academy of Sciences, 1124,* 145–160. doi:10.1196/annals.1440.007

Aharoni, E., Sinnott-Armstrong, W., & Kiehl, K. A. (in press). Can psychopathic offenders discern moral wrongs? A new look at the moral/conventional distinction. *Journal of Abnormal Psychology.*

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin, 126,* 556–574. doi:10.1037/0033-2909.126.4.556

Baird, J. A., & Astington, J. W. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New Directions for Child and Adolescent Development, 103,* 37–49. doi:10.1002/cd.96

Bartels, D. M., & Pizarro, D. (in press). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition.*

Baskin-Sommers, A., Vitale, J., MacCoon, D., & Newman, J. P. (in press). Assessing emotion sensitivity in female offenders with borderline personality features: Results from a fear-potentiated startle paradigm. *Journal of Abnormal Psychology.*

Blair, J., Sellars, C., Strickland, I., Clark, F., Williams, A., Smith, M., & Jones, L. (1996). Theory of Mind in the psychopath. *Journal of Forensic Psychiatry, 7,* 15–25.

Blair, R. J. R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition, 57,* 1–29. doi:10.1016/0010-0277(95)00676-P

Blair, R. J. R. (1997). Moral reasoning and the child with psychopathic tendencies. *Personality and Individual Differences, 22,* 731–739. doi:10.1016/S0191-8869(96)00249-8

Bloom, P. (2011). Family, community, trolley problems, and the crisis in moral psychology. *The Yale Review, 99,* 26–43. doi:10.1111/j.1467-9736.2011.00701.x

Buckholtz, J. W., Treadway, M. T., Cowan, R. L., Woodward, N. D., Benning, S. D., Li, R., . . . Zald, D. H. (2010). Mesolimbic dopamine reward system hypersensitivity in individuals with psychopathic traits. *Nature Neuroscience, 13,* 419–421. doi:10.1038/nn.2510

Cima, M., Tonnaer, F., & Hauser, M. D. (2010). Psychopaths know right from wrong but don't care. *Social Cognitive and Affective Neuroscience, 5,* 59–67. doi:10.1093/scan/nsp051

Cleckley, H. (1941). *The mask of sanity: An attempt to reinterpret the so-called psychopathic personality.* St. Louis, MO: Mosby.

moral judgment. *Cognition, 84,* 221–236. doi:10.1016/S0010-0277 (02)00048-3

O'Kane, A., Fawcett, D., & Blackburn, R. (1996). Psychopathy and moral reasoning: Comparison of two classifications. *Personality and Individual Differences, 20,* 505–514. doi:10.1016/0191-8869(95)00203-0

Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Thinking of mental and other representations: The roles of left and right temporo-parietal junction. *Social Neuroscience, 1*(3–4):245–258. doi:10.1080/17470910600989896

Prinz, J. J. (2004). *Gut reactions.* New York, NY: Oxford University Press.

Prinz, J. J. (2008). *The emotional construction of morals.* New York, NY: Oxford University Press.

Richell, R. A., Mitchell, D. G., Newman, C., Leonard, A., Baron-Cohen, S., & Blair, R. J. (2003). Theory of mind and psychopathy: Can psychopathic individuals read the 'language of the eyes'? *Neuropsychologia, 41,* 523–526. doi:10.1016/S0028-3932(02)00175-6

Royzman, E. B., & Kumar, R. (2004). Is consequential luck morally inconsequential? Empirical psychology and the reassessment of moral luck. *Ratio, 17,* 329–344. doi:10.1111/j.0034-0006.2004.00257.x

Sargent, M. J. (2004). Less thought, more punishment: Need for cognition predicts support for punitive responses to crime. *Personality and Social Psychology Bulletin, 30,* 1485–1493. doi:10.1177/0146167204264481

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind." *NeuroImage, 19,* 1835–1842. doi:10.1016/S1053-8119(03)00230-1

Schaich-Borg, J., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An FMRI investigation. *Journal of Cognitive Neuroscience, 18,* 803–817. doi:10.1162/jocn.2006.18.5.803

Simon, B., Holzberg, J. D., & Unger, J. F. (1951). A study of judgment in the psychopathic personality. *Psychiatric Quarterly, 25,* 132–150. doi:10.1007/BF01584268

Skeem, J., Johansson, P., Andershed, H., Kerr, M., & Eno Louden, J. (2007). Two subtypes of psychopathic violent offenders that parallel primary and secondary variants. *Journal of Abnormal Psychology, 116,* 395–409. doi:10.1037/0021-843X.116.2.395

Trevethan, S., & Walker, L. (1989). Hypothetical versus real-life moral reasoning among psychopathic and delinquent youth. *Development and Psychopathology, 1,* 91–103. doi:10.1017/S0954579400000286

Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context

shape moral judgment. *Psychological Science, 17,* 476–477. doi: 10.1111/j.1467-9280.2006.01731.x

Wertz, A. E., & German, T. C. (2007). Belief-desire reasoning in the explanation of behavior: Do actions speak louder than words? *Cognition, 105,* 184–194. doi:10.1016/j.cognition.2006.08.002

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science, 16,* 780–784. doi:10.1111/j.1467-9280.2005.01614.x

Young, L., Bechara, A., Tranel, D., Damasio, H., Hauser, M., & Damasio, A. (2010). Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. *Neuron, 65,* 845–851. doi:10.1016/j.neuron .2010.03.003

Young, L., Camprodon, J., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporo-parietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgment. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 107,* 6753–6758. doi:10.1073/pnas.0914826107

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences, 104,* 8235–8240. doi:10.1073/ pnas.0701408104

Young, L., Nichols, S., & Saxe, R. (2010). Investigating the neural and cognitive basis of moral luck: It's not what you do but what you know. *Review of Philosophy and Psychology, 1,* 333–349. doi:10.1007/s13164-010-0027-y

Young, L., & Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *NeuroImage, 40,* 1912–1920. doi: 10.1016/j.neuroimage.2008.01.057

Young, L., & Saxe, R. (2009a). An FMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience, 21,* 1396–1405. doi:10.1162/jocn.2009.21137

Young, L., & Saxe, R. (2009b). Innocent intentions: A correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia, 47,* 2065–2072. doi:10.1016/j.neuropsychologia.2009.03.020

Zachary, R. A. (1986). *Shipley Institute of Living Scale: Revised Manual.* Los Angeles, CA: Western Psychological Services.