ELSEVIER

Evolution and Human Behavior 32 (2011) 204 – 215

Evolution and Human Behavior

Original Article

The omission effect in moral cognition: toward a functional explanation

Peter DeScioli[a,*], Rebecca Bruening[b], Robert Kurzban[b]

[a]Chapman University, Economic Science Institute, One University Drive, Orange, CA, USA
[b]University of Pennsylvania, USA

Initial receipt 20 May 2010; final revision received 17 January 2011

Abstract

Moral judgment involves much more than computations of the expected consequences of behavior. A prime example of the complexity of moral thinking is the frequently replicated finding that violations by omission are judged less morally wrong than violations by commission, holding intentions constant. Here we test a novel hypothesis: Omissions are judged less harshly because they produce little material evidence of wrongdoing. Evidence is crucial because moral accusations are potentially very costly unless supported by others. In our experiments, the omission effect was eliminated when physical evidence showed that an omission was chosen. Perpetrators who "opted out" by pressing a button that would clearly have no causal effects on the victim, rather than rescuing them, were judged as harshly as perpetrators who directly caused death. These results show that, to reduce condemnation, omissions must not only be noncausal, they must also leave little or no material evidence that a choice was made.
© 2011 Elsevier Inc. All rights reserved.

Keywords: Omission; Transparency; Moral judgment; Moral psychology

The difference in moral value between omissions and actions is acknowledged by religions, discussed by philosophers and encoded in laws. Hindus, for example, believe that it is morally wrong to kill a cow, but less wrong to let a cow starve to death. India is inundated with hundreds of thousands of stray cattle because people abandon rather than slaughter unwanted animals (Fox, 2003). What explains the widespread moral intuition that a violation by omission is less wrong than a violation by commission?

We approach this question from two perspectives. First, we follow research in moral psychology that takes a cognitive approach by considering the representations and rules that underlie moral judgment (Darley & Pittman, 2003; Gray & Wegner, 2009; Hauser, 2006; Knobe, 2005; Mikhail, 2007; Pizarro, Uhlmann, & Bloom, 2003). Second, we take the perspective that moral cognition is a set of evolved cognitive adaptations which are functionally organized to solve problems associated with moral violations (Alexander, 1987; Darwin, 1871; de Waal, 1996; Haidt, 2007; Lieberman, Tooby, & Cosmides, 2003; 2007; Miller, 2007; Ridley, 1996; Wright, 1994). These two approaches lead us to ask whether patterns in moral judgment, such as the omission effect, provide insight into the functions of the underlying cognitive systems and, conversely, whether functional hypotheses can help uncover new properties of moral judgment.

1. Introduction

1.1. The omission effect

The omission effect is a robust phenomenon that is easily replicable in the lab (Anderson, 2003; Baron & Ritov, 2004; Cushman, Young, & Hauser, 2006; Hauser, 2006). For example, people judge that it is very wrong to poison someone but less wrong to withhold the antidote from someone who has been poisoned, even though the intended consequences are the same (Cushman et al., 2006). Moreover, the effect is implicit in many studies of moral dilemmas in which individuals can choose inaction (e.g., omission vs. killing one person to save others; Mikhail, 2007; Waldmann & Dieterich, 2007). Previous research on omissions spans moral (Cushman et al., 2006; Haidt & Baron, 1996; Kordes-de Vaal, 1996) and nonmoral (Kahneman & Miller, 1986) decision-making. Previous work can also be divided by whether the focus is on omission decisions (e.g., Ritov & Baron, 1999) or on third-party judgments of

* Corresponding author.
E-mail address: pdescioli@yahoo.com (P. DeScioli).

1090-5138/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.evolhumbehav.2011.01.003

others' omissions (e.g., Cushman et al., 2006). It remains an open question how these two phenomena are related (Anderson, 2003; DeScioli, Christner, & Kurzban, in press).

Several explanations for the omission effect have been offered. One theory is that the effect derives from differences in physical causality (Baron & Ritov, 2004). Omissions do not have immediate mechanical effects, whereas actions are clearly causal. Theorists generally regard causality as a necessary condition for blame (Heider, 1958; Shaver, 1985; Weiner, 1995). Summarizing this literature, Alicke (1992) wrote that "causal participation is the basic precondition for ascribing blame and responsibility in virtually all attributional theories of responsibility" (p. 368).

Another set of ideas surrounds anticipated regret (Anderson, 2003; Baron & Ritov, 1994; Kahneman & Miller, 1986). On this account, it is easier to imagine counterfactual outcomes for actions than for omissions; consequently, emotions are amplified for actions. Reduced regret for omissions could explain why people tend to choose violating omissions rather than violating actions. Note, however, that this idea does not necessarily generalize to explain why people condemn *others'* omissions less harshly than others' commissions, i.e., omission effects in third-party judgment (Anderson, 2003).

Here we focus on third-party judgment of omissions. An advantage of this approach is that an explanation for omission effects in third-party judgment simultaneously provides a potential explanation for why actors choose omissions — anticipated blame (Anderson, 2003; DeScioli, Christner, & Kurzban, in press). That is, if third parties judge omissions less harshly, then actors can choose omissions to incur less blame.

1.2. Third-party coordination and public evidence

Why might third parties view omissions with less moral hostility? To try to understand this, we adopt a recently developed theoretical framework which focuses on the problems faced by third parties to moral violations (DeScioli & Kurzban, 2009). This framework takes the perspective that moral cognition is a set of evolved cognitive adaptations which are functionally organized to solve problems associated with moral violations (Alexander, 1987; Darwin, 1871; de Waal, 1996; Haidt, 2007; Hauser, 2006; Lieberman, Tooby, & Cosmides, 2003; 2007; Miller, 2007; Ridley, 1996; Wright, 1994). Importantly, the problems associated with moral events differ according to individuals' roles in the situation such as whether they are a perpetrator, victim or third-party condemner (DeScioli & Kurzban, 2009). One crucial problem faced by third parties is coordinating their condemnation decisions with other third parties (DeScioli, 2008). Here we consider the idea that it is more difficult to coordinate condemnation for omissions than for commissions.

Third-party condemnation and punishment, whether aimed at increasing group welfare (Boyd & Richerson, 2005; Fehr, Fischbacher, & Gächter, 2002) or improving one's reputation (Gintis, Smith, & Bowles, 2001; Kurzban, DeScioli, & O'Brien, 2007), are costly to perform. Condemnation often provokes retaliation from the target and their allies (Knauft, 1987; Miller, 2003; Nikiforakis, 2008; Wiessner, 2005). A morally motivated attack is as risky as a nonmoral attack and should be deployed with equal caution.

If condemners band together, however, the costs of condemnation can be defrayed. Perpetrators can effectively retaliate against only a limited number of people, so condemners can minimize their individual costs by teaming up against perpetrators. This implies that the cost of condemnation varies with the number of condemners. At the extremes, lone accusers face maximum costs, whereas accusers with unanimous support incur minimum costs. Therefore, a well-designed cognitive system for condemnation should evaluate the likelihood that others will condemn, using this information to estimate the costs of moral aggression. All else equal, when less condemnation is expected from others, these cognitive systems should reduce moral hostility.

The coordination problem among condemners raises questions about how individuals predict others' condemnation behavior. One important factor is the public evidence available to other third parties. When strong evidence implicates the accused, others should be more likely to condemn than when the evidence is weak. Furthermore, the availability of good evidence might have another derivative effect: Other third parties will be more likely to condemn not only because of the evidence but because they know that other third parties know the evidence (and so on through infinite recursions). In game theoretic terms, public evidence provides *common knowledge* to third parties about a moral violation, and common knowledge is critical for solving coordination problems (Schelling, 1960).

One difference between wrongful omissions and commissions is that it tends to be more difficult to provide evidence for violating omissions. First, actions have mechanical effects which leave physical evidence such as footprints and fingerprints, whereas omissions are characterized by the *absence* of an action and its mechanical effects; therefore, omissions are less likely to leave physical traces. Second, omissions provide less evidence about the intentions of the actor. Omission is intended if the actor chooses inaction, but omission could also be unintended if the actor is unaware of the circumstances. Consequently, third parties will tend to be more uncertain about intentions for omissions, and given the importance of intentions for blame (e.g., Shaver, 1985; Weiner, 1995), third parties should condemn omissions less harshly.

Crucially, however, even when third parties are confident about wrongful intentions, it might be difficult to convince others that an omission was intended. Further, because other third parties are also attempting to minimize condemnation costs, any residual uncertainty can interrupt common knowledge and coordination. We will refer to the strength of evidence for a violation as its public "transparency"

vs. "opacity," distinguishing this variable from a witness' private confidence about whether a violation occurred. Because the costs of condemnation vary with the number of condemners, it is the public rather than private information about the violation that is most relevant for computing condemnation costs. In short, third parties can coordinate by condemning transparent violations for which there is public evidence while reducing moral hostility toward opaque violations.

These ideas closely parallel a recent game theoretic account of indirect speech (Pinker, Nowak, & Lee, 2008). By using indirect speech for bribes, threats, requests and sexual invitations, individuals can make their intentions deniable. Similarly, third parties might be more reluctant to condemn an omission because it is, effectively, "indirect behavior."

1.3. The present experiments

Based on the theoretical framework summarized above, we hypothesize that people condemn omissions less harshly than commissions because omissions produce little material evidence of wrongdoing. This model straightforwardly predicts that when public evidence shows that an omission was chosen, the omission effect should be reduced or eliminated.

Canonical omissions, relative to actions, tend to simultaneously differ in causality, ambiguity about intentions and public transparency. Our experiments were designed to disentangle these factors by independently varying causality and transparency, holding certainty about intentions constant. This was facilitated by the introduction of a "do-nothing" button, which had no causal effects on the violating event but increased the transparency of the perpetrator's decision process. When an actor "opted out" by pressing a button that would clearly have no effect on a victim's impending death (rather than rescuing them), the material evidence showed that a choice was made to allow the victim's death.

We designed scenarios in which an actor made a choice that was associated with someone's death, and participants judged the actor's behavior. Our experimental designs did not involve moral dilemmas (with ambiguity about which decision is morally best) but simply depicted wrongful behavior and asked participants to judge the magnitude of the offense.

Causality-based theories predict that whether actors physically caused the victim's death should be the key variable that explains differences in perceptions of violation severity. In contrast, the transparency hypothesis predicts that the evidence for an offense influences condemnation severity. Thus, increasing transparency should amplify condemnation even when intentions are held constant and physical causality is absent. That is, the transparency model predicts an interaction in which removing causality reduces condemnation when evidence is unavailable (opaque), but the effect of removing causality is reduced or eliminated when public evidence shows that an omission was chosen (transparent).

2. Experiment 1

2.1. Method

2.1.1. Design

Participants read short scenarios about an individual whose behavior was associated with a victim's death (Appendix A). The scenarios described observable events and avoided assertions of unobservable mental states (e.g., the actor "sees" rather than "knows"). The first scenario involved the demolition of a building (Figure S1). A sequential demolition of three buildings was scheduled to begin in 10 s when an unaware victim arrived on the scene. An actor had access to a computerized control board with buttons that could kill or rescue the victim by changing the order of detonations. The second scenario involved an approaching train (Figure S2). A train approached a station where it could be diverted onto two sidetracks. An actor had access to a computerized control board that could kill or rescue the victim by diverting the train.

Four versions of each scenario were created to vary the causality (none vs. direct) and transparency (transparent vs. opaque) of the actor's behavior. In the no-causality conditions, actors did not cause the event that killed the person. To manipulate transparency, the person either "timed out" by doing nothing or "opted out" by pressing a button that would clearly have no effect on the killing event. In the train scenarios, David saw that he could help but did nothing, whereas Charles pressed a "Maintain Route" button that "has no effect at all on the train, but … updates the computer on the location of the train." In both cases, no causal effect occurred, but opting out was more transparent because it involved movement and left physical evidence (recorded by a computer).

In the direct causality conditions, actors physically caused the event that killed the person. To manipulate transparency, we used an alternative motive. An alternative motive might create opacity because the person can claim that they acted for another reason without knowledge of the victim. In the transparent version, the actors altered the course of events to kill the victim. In the opaque version, they altered the course of events for a different reason (e.g., protecting a bicycle) even though they saw that this would kill the person.

Scenarios were designed to control for intent by making it clear that, in all cases, the victim's death was foreseen and intentional. This is important to distinguish public transparency from uncertainty about intent. Although using alternative motives did alter actors' goals, these alternative goals were designed to be trivial (e.g., protecting a bicycle) to minimize the effect on perceived intent.

For each scenario, participants answered questions about one focal individual. Participants rated moral wrongness and indicated how much prison time the focal individual deserved.

2.1.2. Participants, materials and procedure

Participants were 135 undergraduates (59 males, 76 females) enrolled in introductory psychology at the University

of Pennsylvania. The mean (S.D.) age of our sample was 19.56 (1.75) years.

The experiment was conducted with pencil and paper in the Penn Laboratory for Experimental Evolutionary Psychology at the University of Pennsylvania. Each participant responded to both demolition and train scenarios. They were randomly assigned to one causality condition and one transparency condition. For each scenario, participants answered the following items.

2.1.2.1. Moral wrongness scale. Participants answered "How morally wrong is this behavior?" by rating wrongness from 0 (*not wrong at all*) to 100 (*most wrong*). We provided participants with six scale anchors taken from the results of the National Survey of Crime Severity (Wolfgang, Figlio, Tracey, & Singer, 1985): noise disturbance ("1"), assault with lead pipe causing injury ("10"), knife stabbing causing injury ("20"), knife stabbing causing death ("35"), rape resulting in death ("50") and planting a bomb causing 20 deaths ("70").

2.1.2.2. Deserved punishment. Participants answered "What punishment does this behavior deserve?" by assigning prison time up to 50 years. We provided participants with six anchors taken from the 2006 US Guidelines for Prison Sentencing: trespassing (0–6 months), firearm possession (10–16 months), robbery (33–41 months), sexual abuse (97–121 months), rape (151–188 months) and espionage (360 months–life).

2.1.2.3. Wrongness and punishment comparisons. After participants rated wrongness and punishment, they made forced-choice comparisons between actors in transparent and opaque situations. Participants read two scenarios and decided which actor was more wrong and which deserved more punishment.

After completing the experiment, participants provided demographic information and then they were debriefed and dismissed. The procedure took 30 min. Procedures were approved by the University of Pennsylvania Institutional Review Board.

2.2. Results

2.2.1. Moral wrongness: demolition

Fig. 1 shows wrongness ratings by transparency and causality. Wrongness ratings were analyzed with a 2 (transparency) × 2 (causality) analysis of variance (ANOVA). The transparency × causality interaction was significant, $F(1, 131)=9.15$, $p<.01$. We performed planned comparisons to examine the experimental hypotheses.

2.2.1.1. Transparency effects. When causality was direct, transparent ($M=58.48$) and opaque conditions ($M=59.70$) did not differ, $F(1, 131)=0.03$, $p=.85$. However, perpetrators who had no causal effect on killing were viewed as more wrong in the transparent condition ($M=54.86$) than in the opaque condition ($M=28.55$), $F(1, 131)=17.09$, $p<.001$.

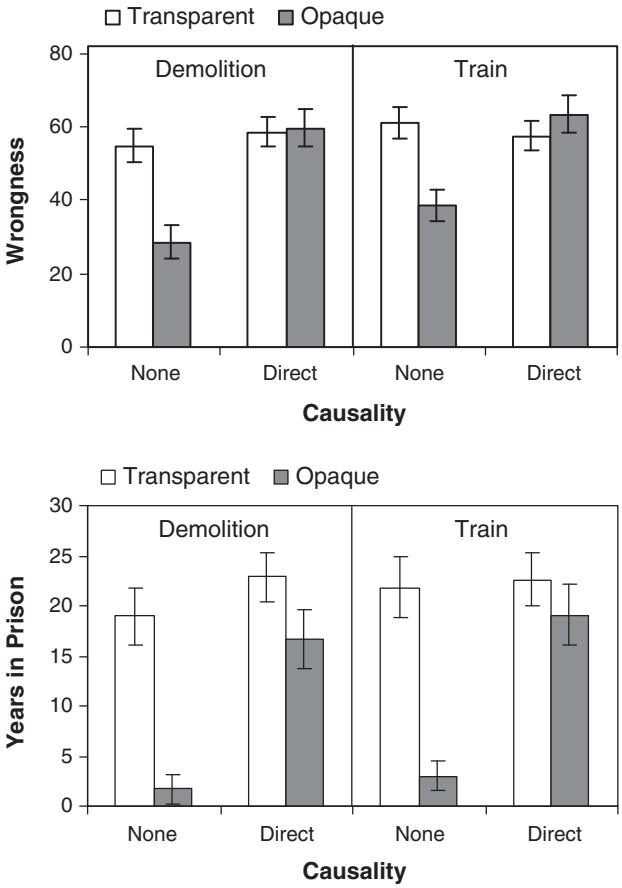

Fig. 1. Mean (S.E.) wrongness and punishment judgments in experiment 1 by transparency, causality and scenario. Wrongness was rated from 0 (*not wrong at all*) to 100 (*most wrong*). Participants assigned a prison sentence from 0 to 50 years. Sample sizes in the no-causality conditions were $n=36$ and 33 for transparent and opaque, respectively; in the direct causality conditions, these values were $n=33$ and 33.

2.2.1.2. Causality effects. In the transparent condition, none ($M=54.86$) and direct ($M=58.48$) were not significantly different, $F(1, 131)=0.32$, $p=.57$. That is, when behavior was transparent in both situations, there was no significant difference between a perpetrator who had no physical causal effect and a perpetrator who directly caused death. In the opaque condition, none ($M=28.55$) was rated as less wrong than direct ($M=59.70$), $F(1, 131)=22.95$, $p<.001$.

2.2.2. Moral wrongness: train

Wrongness ratings were analyzed with a 2 (transparency) × 2 (causality) ANOVA. The two-way transparency × causality interaction was significant, $F(1, 131)=9.94$, $p<.01$. We performed planned comparisons to examine the experimental hypotheses.

2.2.2.1. Transparency effects. When causality was direct, transparent (M=57.42) and opaque conditions (M=63.45) did not differ, $F(1, 131)=0.86$, $p=.35$. However, when perpetrators had no causal effect on killing, they were viewed as more wrong in the transparent condition (M=61.22) than in the opaque condition (M=38.61), $F(1, 131)=12.66$, $p<.001$.

2.2.2.2. Causality effects. In transparent conditions, none (M=61.22) and direct (M=57.42) were not significantly different, $F(1, 131)=0.36$, $p=.55$. Like for demolition, causality showed no effects when behavior was transparent in both situations. In opaque conditions, none (M=38.61) was rated as less wrong than direct (M=63.45), $F(1, 131)=14.64$, $p<.001$.

2.2.3. Punishment: demolition

Fig. 1 shows punishment (years in prison) by transparency and causality. Punishment was analyzed with a 2 (transparency) × 2 (causality) ANOVA. The transparency × causality interaction was significant, $F(1, 131)=4.71$, $p<.05$. We performed planned comparisons to examine the experimental hypotheses.

2.2.3.1. Transparency effects. When causality was direct, prison time assignments for transparent (M=22.86 years) and opaque conditions (M=16.66 years) did not differ, $F(1, 131)=2.93$, $p=.09$. However, perpetrators who had no causal effect on killing were assigned longer prison sentences in the transparent condition (M=18.93 years) than in the opaque condition (M=1.72 years), $F(1, 131)=23.51$, $p<.001$.

2.2.3.2. Causality effects. In transparent conditions, none (M=18.93) and direct (M=22.86 years) were not significantly different, $F(1, 131)=1.22$, $p=.27$. In opaque conditions, none (M=1.72 years) received less prison time than direct (M=16.66 years), $F(1, 131)=16.96$, $p<.001$.

2.2.4. Punishment: train

Punishment was analyzed with a 2 (transparency) × 2 (causality) ANOVA. The transparency × causality interaction was significant, $F(1, 131)=8.28$, $p<.01$. We performed planned comparisons to examine the experimental hypotheses.

2.2.4.1. Transparency effects. When causality was direct, prison time assignments for transparent (M=22.59 years) and opaque conditions (M=19.08 years) did not differ, $F(1, 131)=0.85$, $p=.36$. However, perpetrators who had no causal effect on killing were assigned longer prison sentences in the transparent condition (M=21.86 years) than in the opaque condition (M=3.02 years), $F(1, 131)=25.54$, $p<.001$.

2.2.4.2. Causality effects. In transparent conditions, none (M=21.86 years) and direct (M=22.59 years) were not significantly different, $F(1, 131)=0.04$, $p=.85$. In opaque conditions, none (M=3.02 years) received less prison time than direct (M=19.08 years), $F(1, 131)=17.80$, $p<.001$.

2.2.5. Transparency comparisons

When causality was direct, more participants viewed transparent as morally worse than opaque (demolition: 76%, $p<.001$, binomial test; train: 65%, $p<.05$). With no causality, opting out was viewed as more wrong than doing nothing (demolition: 91%, $p<.001$; train: 97%, $p<.001$).

Turning to punishment comparisons, when causality was direct, transparent offenses were judged to deserve more punishment in the demolition scenario (74%, $p<.001$) but not the train scenario (62%, $p=.06$). With no causality, opting out was viewed as deserving more punishment than doing nothing (demolition: 96%, $p<.001$; train: 99%, $p<.001$).

2.2.6. Summary of results

When perpetrators directly caused the victim's death, the mitigating effect of a trivial alternative motive was modest, visible only in forced-choice comparisons. When Bart caused the man to die because he wanted to watch the train go by, he was judged nearly the same as Alan, who diverted the train for no discernable reason aside from killing the victim. It is possible that this transparency manipulation was too weak or, alternatively, that physical causality alone provides sufficient public evidence.

However, we observed robust transparency effects in the no-causality conditions for both wrongness and punishment judgments. Individuals who opted out by pressing a "do nothing" button were condemned much more harshly than those who did nothing. This difference occurred despite the fact that, in both situations, the perpetrators had no causal effect on the killing and they could have saved victim. In the demolition condition, for example, individuals who did nothing were judged to deserve an average of just 2 years in prison, whereas individuals who opted out by pressing a do-nothing button were sentenced to 19 years.

Causality by itself had surprisingly little influence on condemnation. Strikingly, we observed no difference between opting out (with no causal effect) and directly causing a victim's death. This was true for both wrongness and punishment judgments in both demolition and train scenarios. When behavior was transparent, causality showed no effects.

3. Experiment 2

In experiment 1, we attempted to hold intent constant by making it clear in the scenarios that actors who did nothing knew that they could rescue the victim. However, it is possible that participants nonetheless felt more uncertain about intentions for actors who did nothing. If so, the striking difference between doing nothing and pressing a do-nothing button might be due to uncertainty about intentions, rather than transparency.

We tested this possibility in experiment 2 by introducing a "thinking aloud" manipulation. In the novel treatment, actors

stated "I could save you, but I'm not going to" before doing nothing or before pressing a do-nothing button. Consistent with ideas about the importance of intentions, we predicted a main effect for stated intentions. More importantly, the transparency hypothesis predicts no interaction — the transparency effect will remain even when intentions are clearly stated.

We also added a postexperiment questionnaire to probe participants' perceptions of the intentions and causal effects of the perpetrators in the scenarios. Our primary interest was to check whether participants correctly understood the scenarios. Also, we were interested in whether participants' moral judgments would influence their perceptions of intentions and causality. This interest stemmed from an accumulating literature showing that moral judgment can influence perceptions of an actor's intentions (Knobe, 2005), their causal effects (Alicke, 1992) and their welfare effects (Haidt, 2001).

3.1. Method

3.1.1. Design

The experiment manipulated transparency and stated/unstated intention while holding physical causality constant (actors had no causal effects). Participants read demolition and train scenarios that were similar to experiment 1 (Appendix B). The unstated condition was identical to the no-cause scenarios from experiment 1. In the stated condition, actors thought aloud, stating their intention not to rescue the victim. Specifically, the actor stated "I could save you, but I'm not going to."

3.1.2. Participants, materials and procedure

Participants were 107 undergraduates (54 males, 53 females) enrolled in introductory psychology at the University of Pennsylvania. The mean (S.D.) age of our sample was 19.83 (2.94) years.

We used the same dependent measures and procedure as in experiment 1. Additionally, we added four postexperiment items to check participants' understanding of the scenarios. Participants indicated agreement on a 7-point scale (1=*disagree completely*, 7=*agree completely*) about whether the actor intended the victim to die, caused the victim to die, understood that the victim was about to die and understood that there was an option to prevent the victim's death.

3.2. Results

3.2.1. Moral wrongness

Fig. 2 shows wrongness ratings by condition. Wrongness for demolition scenarios was analyzed with a 2 (transparency) × 2 (stated vs. unstated) ANOVA. The two-way interaction did not reach significance, $F(1, 103)=3.51$, $p=.064$. There were main effects of transparency, $F(1, 103)=23.71$, $p<.001$, and thinking aloud, $F(1, 103)=13.28$, $p<.001$. As predicted, opting out was rated more wrong than doing nothing in the unstated conditions

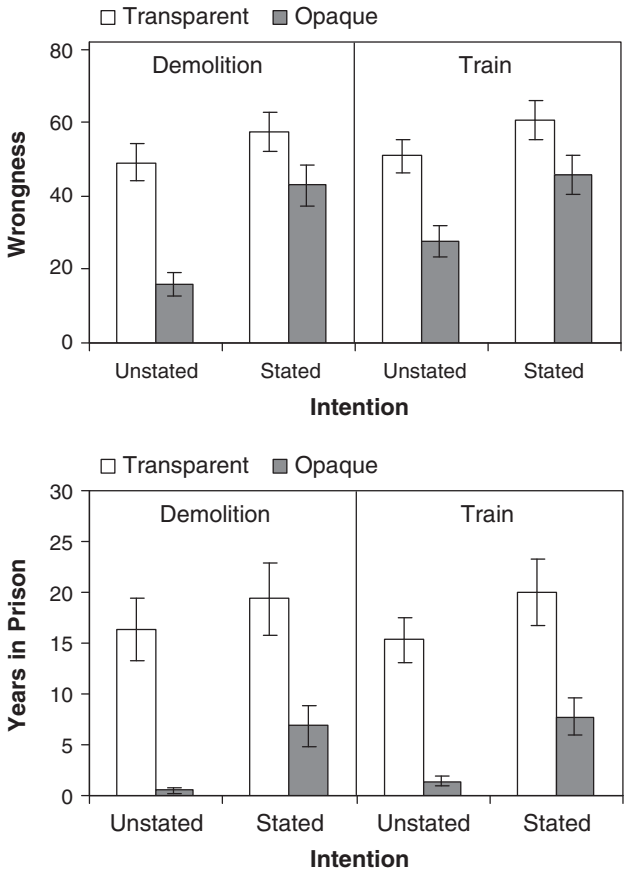

Fig. 2. Mean (S.E.) wrongness and punishment judgments in experiment 2 by transparency, stated/unstated intention and scenario. Wrongness was rated from 0 (*not wrong at all*) to 100 (*most wrong*). Participants assigned a prison sentence from 0 to 50 years. Sample sizes in the unstated conditions were $n=27$ and 25 for transparent and opaque, respectively; in the stated conditions, these values were $n=28$ and 27.

($M=49.00$ vs. $M=15.96$) and in the stated conditions ($M=57.68$ vs. $M=43.00$).

Wrongness for train scenarios was analyzed with a 2 (transparency) × 2 (stated/unstated) ANOVA. The two-way interaction did not reach significance, $F(1, 103)=0.75$, $p=.39$. There were main effects of transparency, $F(1, 103)=14.62$, $p<.001$, and thinking aloud, $F(1, 103)=7.82$, $p<.01$. As predicted, opting out was rated more wrong than doing nothing in the unstated conditions ($M=50.93$ vs. $M=27.60$) and in the stated conditions ($M=60.54$ vs. $M=45.81$).

3.2.2. Punishment

Fig. 2 shows punishment by condition. Punishment for demolition scenarios was analyzed with a 2 (transparency) × 2 (stated vs. unstated) ANOVA. The two-way interaction did not reach significance, $F(1, 103)=0.40$, $p=.53$. There was a main effect of transparency, $F(1, 103)=29.08$, $p<.001$. As predicted, opting out was viewed as deserving more punishment than doing nothing in the unstated conditions

(M=16.39 years vs. M=0.55 years) and in the stated conditions (M=19.37 years vs. M=6.84 years). However, stated intentions were not punished significantly more than unstated intentions, $F(1, 103)=3.11$, $p=.08$.

Punishment for train scenarios was analyzed with a 2 (transparency) × 2 (stated/unstated) ANOVA. The two-way interaction did not reach significance, $F(1, 103)=0.15$, $p=.70$. There were main effects of transparency, $F(1, 103)=33.80$, $p<.001$, and stated intentions, $F(1, 103)=6.04$, $p<.05$. As predicted, opting out was viewed as deserving more punishment than doing nothing in the unstated conditions (M=15.30 years vs. M=1.41 years) and in the stated conditions (M=19.94 years vs. M=7.79 years).

3.2.3. Comparisons

Participants nearly unanimously chose transparent offenses as more wrong and punishable than opaque violations for both scenarios and in stated and unstated conditions (all percentages ≥96%, all ps<.001, binomial test).

3.2.4. Postexperiment questions

Table 1 reports the mean agreement ratings for the postexperiment items. For transparent violations (in both stated- and unstated-intention conditions), participants generally agreed that the actor intended the victim's death, caused the victim's death, understood the danger to the victim and understood that they could save the victim (all Ms>5.00). For opaque violations, the results differed depending on whether the perpetrator's intentions were stated or unstated. When intentions were unstated, participants tended toward disagreement or neutrality with the response items. We suggest caution in interpreting these results, given that previous research has shown that moral judgment can influence judgments about an actor's mental states (Knobe, 2005). As expected, however, when intentions were clearly stated, participants generally agreed that perpetrators who did nothing intended the victim's death, understood the danger and understood that they could save the victim.

Table 1
Mean agreement ratings for experiment 2

Item	Unstated		Stated	
	Transparent	Opaque	Transparent	Opaque
Demolition				
Intended	5.67	1.92	6.07	5.00
Caused	5.78	2.40	5.93	3.93
Understood danger	5.48	3.48	6.75	6.56
Understood option to save	6.86	6.59	6.75	6.63
Train				
Intended	6.15	2.16	6.29	4.93
Caused	6.00	3.08	5.82	4.11
Understood danger	6.19	4.20	6.89	6.52
Understood option to save	6.44	4.40	6.86	6.59

Note. Participants' mean agreement on a 7-point scale (1=*disagree completely*, 7=*agree completely*) about whether the actor intended the victim to die, caused the victim to die, understood that the victim was about to die and understood that there was an option to prevent the victim's death.

The results to the postexperiment items were consistent with expectations except for the causality judgments, which are difficult to understand. None of the scenarios in experiment 2 involved perpetrators who physically caused a victim's death, so these responses are somewhat puzzling. Did participants incorrectly understand the scenario, or is this another case of "culpable causation" (Alicke, 1992) in which moral condemnation influences perceptions of causality? We return to this issue in experiment 3.

3.2.5. Summary of results

The results of experiment 2 show that people continue to judge transparent offenses more harshly than opaque violations, even when the perpetrators in both cases clearly state their intentions to refrain from saving the victim.

4. Experiment 3

In experiment 2, some participants responded that the perpetrator caused the victim's death, although physical causality was absent in the scenarios. This could indicate that participants did not understand the scenarios. Alternatively, participants' moral judgments might have influenced their perceptions of causality, as found in previous work (Alicke, 1992). Still another possibility is that participants might have interpreted the term "caused" more broadly than strict physical causality. In experiment 3, we explored this issue by asking participants to complete a set of comprehension questions before making wrongness and punishment judgments.

4.1. Method

4.1.1. Design

Experiment 3 replicated the stated-intention conditions from experiment 2. That is, perpetrators stated "I could save you, but I'm not going to" before doing nothing or pressing a do-nothing button. The key modification was that immediately after reading a scenario, participants answered three comprehension questions. The critical question asked, "If [NAME] had stayed at home that day instead of walking by the demolition site [train station], would the person standing next to Building B [on the Main Track] have been killed?" If participants understood that the victim would have been killed if the actor was absent, then they correctly understood that the actor did not physically cause the victim's death. By placing this question immediately after the scenario, we could distinguish whether participants correctly understood the scenario while making their subsequent judgments. However, this design also draws the participant's attention to the absence of causality, which might diminish moral condemnation.

4.1.2. Participants, materials and procedure

We recruited 124 participants (62 males, 62 females) to participate in an online study for which they received a small payment. The mean (S.D.) age of our sample was 32.99 (11.97) years.

We used the same dependent measures and procedure as in experiment 2, except we left out the direct comparisons and we added a comprehension check. Participants answered three comprehension questions immediately after reading the scenario. The first question asked for the number of buildings or buttons. The second question asked whether the victim could have been saved (yes or no). The third question was the critical causality question which asked whether the victim would have been killed if the actor had been absent from the event (yes or no).

4.2. Results

4.2.1. Comprehension questions

Overall, participants correctly answered the comprehension questions. The first question was answered correctly by nearly all participants in both conditions and scenarios (all percentages ≥95%). The second question was answered correctly by most participants in the demolition scenario (transparent=83%, opaque=84%) and the train scenario (transparent=97%, opaque=94%). The critical comprehension question was the third item dealing with the causal effects of the actor. Most participants correctly answered the causality question in the demolition scenario (transparent=77%, opaque=95%) and the train scenario (transparent=92%, opaque=92%). These results indicate that some participants might have had some difficulty with the demolition scenario in the transparent condition, and we return to this issue below.

4.2.2. Moral wrongness

We did not observe the predicted transparency effects in wrongness judgments. In the demolition scenario, the difference between transparent (M=46.05) and opaque (M=44.05) conditions was not significant, $F(1, 122)=0.20$, $p=.66$. In the train scenario, the difference between transparent (M=46.13) and opaque (M=43.02) conditions was not significant, $F(1, 122)=0.46$, $p=.46$.

4.2.3. Punishment

In the demolition scenario, the difference between transparent and opaque conditions was significant, $F(1, 122)=8.28$, $p=.005$. As predicted, participants assigned greater punishment for transparent offenses than for opaque offenses (M=17.88 years vs. M=10.30 years). In the train scenario, the difference between transparent and opaque conditions was significant, $F(1, 122)=8.21$, $p=.005$. As predicted, participants assigned greater punishment for transparent offense than for opaque offenses (M=17.88 years vs. M=10.58 years).

The comprehension results suggested that some participants had difficulty understanding the demolition scenario in the transparent condition. To check whether misunderstanding influenced the punishment results, we reanalyzed the data in the demolition scenario excluding participants who incorrectly answered the causality question. We found that difference between transparent and opaque conditions remained significant, $F(1, 104)=6.02$, $p=.02$. Among participants who correctly answered the causality question, transparent offenses were assigned more punishment than opaque offenses (M=16.91 years vs. M=9.92 years).

4.2.4. Postexperiment questions

Table 2 shows the mean agreement ratings for the postexperiment items. We observed the same pattern of results as in the stated-intention condition in experiment 2. As in experiment 2, many participants agreed that the actor caused the victim's death. This occurred despite the fact that most participants correctly answered the comprehension question, showing that they understood that if the actor had been absent, then the victim still would have been killed. Hence, it is clear that participants did not understand "caused" in terms of strictly physical causal processes. Future work can aim to clarify precisely how people understand the concept of causation in the context of moral events.

4.2.5. Summary of results

The primary aim of experiment 3 was to examine whether participants understood the noncausal nature of opting out and timing out. We observed correct responses to the key causality question for over 90% of participants in all conditions except the demolition opt-out condition in which 77% answered correctly. The punishment results replicated the findings of the previous experiments: Perpetrators who opted out were punished more harshly than perpetrators who timed out. This result continued to hold when participants who misunderstood the causality of the demolition scenario were excluded from the analysis. We

Table 2
Mean agreement ratings for experiment 3

Item	Transparent	Opaque
Demolition		
Intended	5.97	4.49
Caused	5.11	4.21
Understood danger	6.61	5.98
Understood option to save	6.21	5.60
Train		
Intended	6.25	4.76
Caused	5.51	4.74
Understood danger	6.70	5.94
Understood option to save	6.65	6.10

Note. Participants' mean agreement on a 7-point scale (1=disagree completely, 7=agree completely) about whether the actor intended the victim to die, caused the victim to die, understood that the victim was about to die and understood that there was an option to prevent the victim's death.

observed no difference in wrongness ratings between opt-out and time-out conditions. This raises questions about the robustness of transparency effects on wrongness ratings. However, we note that, in order to accomplish the main goal of this experiment, participants first answered questions about the causal nature of the perpetrators' actions. This procedure inevitably draws attention to the absence of physical causal effects, and this might explain the inconsistency between experiment 3 and the previous experiments.

Additionally, experiment 3 showed that participants responded that perpetrators caused the victim's death even after correctly answering that, if the perpetrator had stayed home, the victim still would have been killed. This finding adds to previous research showing that moral judgment influences perceptions of causality (Alicke, 1992) and, more generally, to work on top-down processing effects in moral cognition, such as post hoc attributions of intentions (Knobe, 2005) and harm (Haidt, 2001).

5. Experiment 4

We predicted the effects in experiments 1–3 based on the idea that actions provide greater evidence of wrongdoing than omissions. However, the previous studies did not include a manipulation check to test for participants' explicit beliefs about the evidence available. Although it is possible that this information is implicit and inaccessible (e.g., Haidt, 2001; Mikhail, 2007), it is also possible that people's assessments of evidence will correlate with their judgments of wrongdoing. In experiment 4, we investigated this issue by asking participants to assess the evidence showing wrongdoing.

5.1. Method

5.1.1. Design

Experiment 4 replicated the stated-intention conditions from experiment 2. The perpetrators stated "I could save that person, but I'm not going to" before doing nothing or pressing a do-nothing button. After judging wrongness and punishment, participants answered two new questions assessing the evidence. The first question asked, "If the individual denied any wrongdoing, how strong would the evidence be against him?" Participants responded on a 7-point-scale (1=*very weak evidence*, 7=*very strong evidence*). The second question asked, "If this individual denied any wrongdoing, how easy would it be to demonstrate his guilt to other people who were not present to see the events?" Participants responded on a 7-point-scale (1=*very difficult to demonstrate*, 7=*very easy to demonstrate*).

5.1.2. Participants, materials and procedure

We recruited 162 participants (96 females, 66 males) to participate in an online study for which they received a small payment. The mean (S.D.) age of our sample was 36.00 (12.26) years. We used the same dependent measures and procedure as in experiment 2, except we added the two items about evidence after the wrongness and punishment items.

5.2. Results

5.2.1. Moral wrongness

In the demolition scenario, wrongness ratings were greater for the transparent condition (M=51.51) than the opaque condition (M=44.82) with marginal significance, $F(1, 160)=2.77$, $p=.09$. In the train scenario, the difference between transparent (M=51.68) and opaque conditions (M=46.68) was not significant, $F(1, 160)=1.71$, $p=.19$.

5.2.2. Punishment

In the demolition scenario, the difference in punishment judgments between transparent and opaque conditions was significant, $F(1, 160)=20.53$, $p<.001$. As predicted, participants assigned greater punishment for transparent offenses than for opaque offenses (M=17.63 years vs. M=8.62 years). In the train scenario, the difference between transparent and opaque conditions was significant, $F(1, 160)=11.55$, $p<.001$. As predicted, participants assigned greater punishment for transparent offenses than for opaque offenses (M=17.70 years vs. M=10.78 years).

5.2.3. Participants' assessments of evidence

The two evidence items were highly correlated in the demolition scenario ($r=.73$) and the train scenario ($r=.87$); hence, we averaged the items to create a composite measure for evidence judgments. In the demolition scenario, participants viewed the evidence as stronger in the transparent condition than the opaque condition, $F(1, 160)=6.65$, $p=.01$. In the train scenario, the difference between conditions in evidence judgments was significant, $F(1, 160)=2.61$, $p=.05$, one-tailed test.

Furthermore, evidence ratings were correlated with wrongness ratings in the demolition scenario ($r=.32$, $p<.001$) and the train scenario ($r=.33$, $p<.001$). Evidence ratings were also correlated with punishment judgments in the demolition scenario ($r=.43$, $p<.001$) and the train scenario ($r=.32$, $p<.001$).

5.2.4. Summary of results

Experiment 4 replicated the punishment results from previous studies, showing nearly a twofold increase in prison time assigned for transparent offenses relative to opaque offenses. For wrongness ratings, however, the results were mixed, showing greater wrongness for transparent offenses in the demolition scenario, but no significant difference in the train scenario. This cautions that the effect of this manipulation on wrongness judgments is relatively small and potentially fragile.

The novel contribution of experiment 4 is that we found that participants' assessments about evidence differed between transparent and opaque conditions, providing a check on the manipulation used in this and the previous studies. Also, we found that participants' assessments of evidence were correlated with their wrongness and

punishment judgments, supporting the hypothesis that moral condemnation is tuned to the strength of the evidence which could be used to demonstrate the violation.

6. Discussion

Our results present challenges to previous theories for the omission effect and point to an additional variable, offense transparency, which might be important. In experiment 1, opting out of rescuing a victim by pressing a do-nothing button was judged more wrong and punishable than doing nothing. In fact, we observed no statistical differences in moral judgments about, on the one hand, perpetrators who pushed a button that had absolutely no relevant causal effect and, on the other, perpetrators who directly caused another person's death. Experiment 2 showed that transparency affected judgments even when perpetrators explicitly stated their intention to refrain from rescuing the victim. In experiment 3, we found that participants understood the noncausal nature of opting out and timing out. Experiment 3 replicated the previous transparency effect for punishment judgments, but we observed no difference in wrongness ratings. Experiment 4 added a manipulation check showing that assessments of evidence differed between transparency conditions, and further, it showed that participants' assessments of evidence correlated with their wrongness and punishment judgments.

The findings of experiment 3 also add to the literature showing that higher-level moral judgments influence lower-level perceptions of causality, intentions and harm (Alicke, 1992; Haidt, 2001; Knobe; 2005) upon which those moral judgments are typically considered to be based. In experiment 3, participants correctly answered that the victim would still have been killed if the perpetrator had stayed home, showing that they understood that the do-nothing-button had no effect, yet the same participants responded that the perpetrator caused the victim's death.

Why are omissions judged less harshly than actions? The present findings challenge the idea that causality explains omission effects. If causality were the key variable, then doing nothing and opting out ought to be equally condemned. Instead, participants more harshly condemned opting out, even statistically indistinguishable from directly causing the victim's death. This striking result leads us to speculate that the influence of causality on moral judgment, more broadly, might be explained in part by the transparency of causal processes.

The reported experiments also do not fit well with theories that turn on anticipated regret and counterfactual reasoning (Anderson, 2003; Kahneman & Miller, 1986). The counterfactual reasoning for opting out is identical to the reasoning for for timing out. Both require imagining that the perpetrator performed an action (pressing a button) that they did not perform.

The present experiments were motivated by the idea that public evidence influences moral judgment. In the train scenarios, pressing a button that changes the path of the train, causing a victim's death, would lead an after-the-fact observer to infer deliberate wrongdoing. A standard omission, in contrast, provides less evidence because of its noncausal nature. If an omitting perpetrator explicitly stated their wrongful intent, then the evidence is stronger, but this statement is easy to deny. However, when the actor presses a do-nothing button recorded by a computer, the evidence shows that a choice was made to allow the victim's death.

The transparency hypothesis was productive in this study in that it led to experimental results that pose serious difficulties for current theories for the omission effect. However, the theory needs to be subjected to further tests in future research. For example, we used a single type of moral offense, murder, and our participants were mostly Americans. However, there is a tremendous variety of moral rules both within and between human cultures. It remains unclear the extent to which transparency operates across different kinds of moral rules and across the diversity of human cultures. Also important, future work should develop and test alternative hypotheses for the findings of this study. More broadly, future research can further explore how the adaptive problems surrounding condemnation can illuminate its information-processing structure (DeScioli & Kurzban, 2009). This approach can clarify how condemners navigate a complex social world in which moral accusations can be both helpful and dangerous.

Supplementary materials related to this article can be found online at doi:10.1016/j.evolhumbehav.2011.01.003.

Acknowledgments

We thank Alex Shaw for research assistance. This research was supported by a predoctoral fellowship to Peter DeScioli from the International Foundation for Research in Experimental Economics (IFREE).

Appendix A. Stimuli for Experiment 1

Demolition scenario

Participants were presented with the following scenario (Figure S1), which remained constant across conditions.

A construction company is scheduled to demolish three adjacent buildings, which they have labeled A, B and C. Each of the buildings is wired with dynamite. Detonations are controlled by a central control board. A countdown timer indicates when the detonations will occur. The buildings are scheduled to explode one at a time every 30 minutes beginning at noon. The order of detonation is controlled by three buttons labeled A, B and C. Above the buttons is a sign that reads "Press the button to make the corresponding building FIRST in line for detonation." When a button is pressed, a computer

system records the decision. [NAME] is walking down the street when he comes across the control board. The control board reads "Next for detonation: Building B," and the countdown timer currently reads "10 seconds."

This was followed by one of the following four endings:

Direct causality, transparent. "ANDREW sees a man standing next to Building C. ANDREW presses button C. Building C explodes and the man is killed."

Direct causality, opaque. "BILL sees that his bicycle is located between Building A and Building B. BILL reaches to press button C to prevent damage to his bicycle. Before doing so, BILL sees a man standing next to Building C. BILL continues, pressing button C. Building C explodes and the man is killed."

No causality, transparent. "CHRIS sees a man standing between Building A and Building B. CHRIS presses button B. Building B explodes and the man is killed."

No causality, opaque. "DAN sees a man standing between Building A and Building B. DAN does nothing. Building B explodes and the man is killed."

Train scenario

Participants were presented with the following scenario (Figure S2), which remained constant across conditions.

There is a control room at a train station. In this room, a set of buttons controls two railroad switches that can redirect a train onto two sidetracks. When a button is pressed, a computer system records the decision and updates system information accordingly. One button reads "Sidetrack A." When this button is pressed, the train switches off its current track onto Sidetrack A (see diagram). The second button reads "Sidetrack B." When this button is pressed, the train switches off its current track onto Sidetrack B. The third button reads "Maintain Route." Pressing this button has no effect at all on the train, but like the other buttons, it updates the computer system on the location of the train. [NAME] walks into the control room and sees a train coming down the tracks. The train is not slowing down like it usually does, but is traveling at full speed (the driver has fallen asleep).

This was followed by one of the following four endings:

Direct causality, transparent. "ALAN sees that there is a person on Sidetrack B. ALAN presses the "Sidetrack B" button. The person is killed."

Direct causality, opaque. "BART wants to watch the train go by on Sidetrack B from the control room so he reaches to press the "Sidetrack B" button. Before doing so, he sees that there is a person on Sidetrack B. Bart continues, pressing the "Sidetrack B" button. The person is killed."

No causality, transparent. "CHARLES sees that there is a person on the Main Track. CHARLES sees that Sidetrack A can direct the train around the person. CHARLES presses the "Maintain Route" button. The person is killed."

No causality, opaque. "DAVID sees that there is a person on the Main Track. DAVID sees that Sidetrack A can direct the train around the person. DAVID does not press any buttons. The person is killed."

Appendix B. Stimuli for Experiment 2

Experiment 2 manipulated transparency (transparency vs. opaque) and intention (unstated vs. stated). The unstated conditions were identical to the no-causality conditions in experiment 1. In the stated conditions, the actor thought aloud, "I could save you, but I'm not going to."

For the demolition scenario, the endings for the stated conditions were as follows:

Stated, transparent. "AARON sees a man standing between Building A and Building B. AARON says aloud, "I could save you, but I'm not going to." AARON presses button B. Building B explodes and the man is killed."

Stated, opaque. "BERT sees a man standing between Building A and Building B. BERT says aloud, "I could save you, but I'm not going to." BERT does not press any button. Building B explodes and the man is killed."

For the train scenario, the endings for the stated conditions were as follows:

Stated, transparent. "ALEX sees that there is a person on the Main Track. ALEX sees that Sidetrack A can direct the train around the person. ALEX says aloud, "I could save you, but I'm not going to." ALEX presses the "Maintain Route" button. The person is killed."

Stated, opaque. "BEN sees that there is a person on the Main Track. BEN sees that Sidetrack A can direct the train around the person. BEN says aloud, "I could save you, but I'm not going to." BEN does not press any button. The person is killed."

References

Alexander, R. A. (1987). *The biology of moral systems.* New York: Aldine de Gruyter.

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology, 63*, 368–378.

Anderson, C. J. (2003). The psychology of doing nothing: forms of decision avoidance result from reason and emotion. *Psychological Bulletin, 129*, 139–167.

Baron, J., & Ritov, I. (1994). Reference points and omission bias. *Organizational Behavior and Human Decision Processes, 59*, 475–498.

Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes, 94*, 74–85.

Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures.* New York: Oxford University Press.

Cushman, F. A., Young, L., & Hauser, M. D. (2006). The role of reasoning and intuition in moral judgments: testing three principles of harm. *Psychological Science, 17*, 1082–1089.

Darley, J. M., & Pittman, T. S. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Review, 7*, 324–336.

Darwin, C. (1871). *Descent of man, and selection in relation to sex.* New York: D. Appleton and Company.

de Waal, F. (1996). *Good natured: the origins of right and wrong in humans and other animals.* Cambridge, MA: Harvard University Press.

DeScioli, P. (2008). *Investigations into the problems of moral cognition. Unpublished doctoral dissertation.* Philadelphia, PA: University of Pennsylvania.

DeScioli P., Christner J., & Kurzban R. (in press). The omission strategy. *Psychological Science.*

DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition, 112*, 281–299.

Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature, 13*, 1–25.

Fox, M. W. (2003). India's sacred cow: her plight and future. In S. J. Armstrong, & R. G. Botzler (Eds.), *The animal ethics reader* (pp. 238−241). New York: Routledge.

Gintis, H., Smith, E. A., & Bowles, S. (2001). Costly signaling and cooperation. *Journal of Theoretical Biology, 213*, 103–119.

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.

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review, 108*, 814–834.

Haidt, J. (2007). The new synthesis in moral psychology. *Science, 316*, 998–1002.

Haidt, J., & Baron, J. (1996). Social roles and the moral judgment of acts and omissions. *European Journal of Social Psychology, 26*, 201–218.

Hauser, M. D. (2006). *Moral minds.* New York: HarperCollins.

Heider, F. (1958). *The psychology of interpersonal relations.* New York: Wiley.

Kahneman, D., & Miller, D. T. (1986). Norm theory: comparing reality to its alternatives. *Psychological Review, 93*, 136–153.

Knauft, B. M. (1987). Reconsidering violence in simple human societies: homicide among the Gebusi of New Guinea. *Current Anthropology, 28*, 457–500.

Knobe, J. (2005). Theory of mind and moral cognition: exploring the connections. *Trends in Cognitive Sciences, 9*, 357–359.

Kordes-de Vaal, J. (1996). Intention and the omission bias: omissions perceived as nondecisions. *Acta Psychologica, 93*, 161–172.

Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior, 28*, 75–84.

Lieberman, D., Tooby, J., & Cosmides, L. (2003). Does morality have a biological basis? An empirical test of the factors governing moral sentiments relating to incest. *Proceedings of the Royal Society B, 270*, 819–826.

Lieberman, D., Tooby, J., & Cosmides, L. (2007). The architecture of human kin detection. *Nature, 445*, 727–731.

Mikhail, J. (2007). Universal moral grammar: theory, evidence and the future. *Trends in Cognitive Sciences, 11*, 143–152.

Miller, G. F. (2007). Sexual selection for moral virtues. *Quarterly Review of Biology, 82*, 97–125.

Miller, G. P. (2003). Norm enforcement in the public sphere: the case of handicapped parking. *George Washington Law Review, 71*, 895–933.

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: can we really govern ourselves? *Journal of Public Economics, 92*, 91–112.

Pinker, S., Nowak, M. A., & Lee, J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences, 105*, 833–838.

Pizarro, D. A., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology, 39*, 653–660.

Ridley, M. (1996). *The origins of virtue.* London: Viking: Penguin Books.

Ritov, I., & Baron, J. (1999). Protected values and omission bias. *Organizational Behavior and Human Decision Processes, 79*, 79–94.

Schelling, T. C. (1960). *The strategy of conflict.* Cambridge, MA: Harvard University Press.

Shaver, K. G. (1985). *The attribution of blame: causality, responsibility, and blameworthiness.* New York: Springer-Verlag.

Waldmann, M. R., & Dieterich, J. (2007). Throwing a bomb on a person versus throwing a person on a bomb: intervention myopia in moral intuitions. *Psychological Science, 18*, 247–253.

Weiner, B. (1995). *Judgments of responsibility: a foundation for a theory of social conduct.* New York: Guilford Press.

Wiessner, P. (2005). Norm enforcement among the Ju/'hoansi Bushmen: a case of strong reciprocity? *Human Nature, 16*, 115–145.

Wolfgang M. E., Figlio R. M., Tracy P. E., & Singer S. I. (1985). *The National Survey of Crime Severity*. Washington, D.C.: U.S. Dept. of Justice, Bureau of Justice Statistics.

Wright, R. (1994). *The moral animal.* New York: Pantheon.

Supporting Information

Figure S1

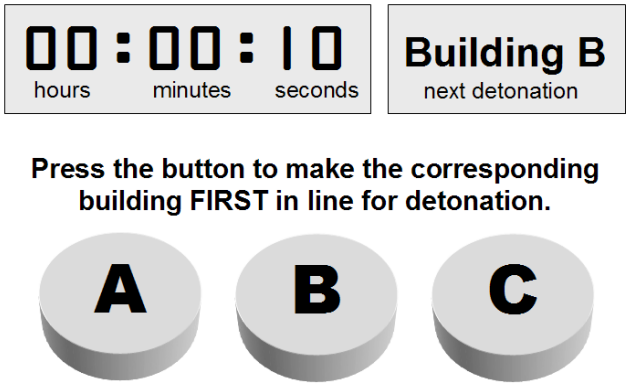

Figure S1. An image from the experimental stimuli for the demolition scenario.

Figure S2

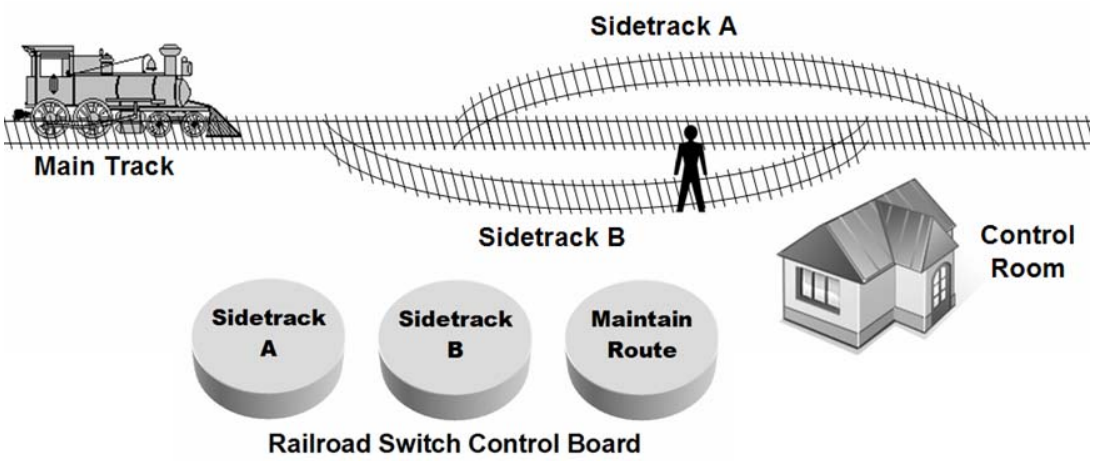

Figure S2. An image from the experimental stimuli for the train scenario.