ELSEVIER

Evolution and Human Behavior xx (2012) xxx – xxx

Evolution and Human Behavior

Original Article

Hamilton vs. Kant: pitting adaptations for altruism against adaptations for moral judgment

Robert Kurzban[a,*], Peter DeScioli[b], Daniel Fein[a]

[a]Department of Psychology, University of Pennsylvania, Philadelphia PA 19104, USA
[b]Departments of Psychology and Economics, Brandeis University, Waltham, MA 02453, USA

Initial receipt 22 February 2010; final revision received 3 November 2011

Abstract

Prominent evolutionary theories of morality maintain that the adaptations that underlie moral judgment and behavior function, at least in part, to deliver benefits (or prevent harm) to others. These explanations are based on the theories of kin selection and reciprocal altruism, and they predict that moral systems are designed to maximize Hamiltonian inclusive fitness. In sharp contrast, however, moral judgment often appears Kantian and rule-based. To reconcile this apparent discrepancy, some theorists have claimed that Kantian moral rules result from mechanisms that implement simple heuristics for maximizing welfare. To test this idea, we conducted a set of studies in which subjects (N=1290) decided whether they would kill one person to save five others, varying the relationship of the subject with the others involved (strangers, friends, brothers). Are participants more likely to observe the Kantian rule against killing in decisions about brothers and friends, rather than strangers? We found the reverse. Subjects reported greater willingness to kill a brother or friend than a stranger (in order to save five others of the same type). These results suggest that the rule-based structure of moral cognition is not explained by kin selection, reciprocity, or other altruism theories.

Keywords: Morality; Altruism; Kin selection; Inclusive fitness; Judgment & decision making; Reasoning

1. Introduction

1.1. Human moral cognition is nonconsequentialist

Consider an organism faced with a dilemma. It can either kill one of its offspring, which will allow five others to live, or it can do nothing, in which case five of its offspring will die. Hamilton's (1964) theory of kin selection predicts that evolution will favor designs for killing one relative to save five others. Indeed, this behavior has been observed in many species, such as the dramatic case of the burying beetle, which kills some offspring in order to feed the bodies to other offspring (Mock, 2004). The burying beetle's decisions are, just as kin selection predicts, *consequentialist*, based exclusively on outcomes. More precisely, the mechanism that causes the burying beetle's infanticide-and-regurgitation behavior was selected by virtue of the inclusive fitness consequences of its choices.

Immanuel Kant would argue, however, that when humans face this dilemma, they should *not* kill one to save five because there is an inviolable moral rule against killing that cannot be broken regardless of the consequences. Kant's view—*nonconsequentialism*—is reflected in a characteristic feature of human moral cognition: Moral judgment is rule-based and focuses on *behavior per se*, the means used to accomplish outcomes, rather than on the *outcomes*, or ends (DeScioli & Kurzban, 2009a). For instance, in the footbridge version of the Trolley Problem (Foot, 1967; see Methods, below), 90% of people judge that it is impermissible to kill one person to save five people (Hauser, Young, & Cushman, 2008).

Why is the burying beetle's behavior consequentialist while human moral judgment is nonconsequentialist? Instead of using simple rules such as "never kill," "never steal," or "never eat pork," humans could make moral decisions based on only the costs and benefits of their options. The phenomenon of nonconsequentialism in moral

* Corresponding author. Department of Psychology, University of Pennsylvania, Philadelphia PA 19104, USA. Tel.: +1 215 898 4977; fax: +1 215 898 7301.
E-mail address: kurzban@psych.upenn.edu (R. Kurzban).

1090-5138/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.evolhumbehav.2011.11.002

judgment is easily overlooked as a puzzle because it is so familiar and intuitive (Cosmides & Tooby, 1994). But this feature poses a problem: Why do humans focus moral decisions on behavior rather than considering only the consequences?

1.2. Describing the problem with choice theory

Basic choice theory clarifies the distinction between consequentialism and nonconsequentialism. In choice theory, there is a decision maker who selects an action, *a,* from a set of possible actions, A, and each action is associated with possible outcomes. Here, an outcome consists of payoffs to the organism and other relevant organisms, i.e., a vector $\mathbf{y}$ of payoffs to the self and others. Finally, the decision maker has a standard utility function for ranking outcomes depending on the resulting payoffs, $u(\mathbf{y})$.

A consequentialist decision procedure would choose an action, a^*, to maximize utility:

$$\max_{a \in A} u(\mathbf{y}), \tag{1}$$

where $u(\mathbf{y})$ depends only on the vector of payoffs, $\mathbf{y}$. This encompasses a range of utility functions including any weighted sum of payoffs to the self and others, whether characterized by extreme selfishness, universal altruism, or altruism skewed toward family and friends.

In contrast, the Kantian decision procedure would choose an action not only based on the payoffs $\mathbf{y}$, but also based on whether the action is labeled morally wrong. Morally wrong actions are excluded regardless of the payoffs they generate. The Kantian approach can be expressed as maximization subject to constraints on the actions:

$$\max_{a \in A} u(\mathbf{y}), \text{subject to the constraint}, a \notin W, \tag{2}$$

where W refers to a set of actions labeled morally wrong. The Kantian decision rule excludes actions in W and then maximizes utility subject to this constraint.

Moral dilemmas occur when maximization based on payoffs (1) conflicts with moral constraints (2). Specifically, a dilemma arises when the action that maximizes utility for the consequentialist, a^*, is in the set W of moral wrongs. In these situations, decision processes (1) and (2) lead to different choices. Empirical observations show that people's choices are sometimes most consistent with (1), as in the switch version of the Trolley Problem, and sometimes with (2), as in the footbridge version of the Trolley Problem (Hauser, 2006). This pattern of results suggests that both of these conflicting decision processes are used to some extent, which is presumably why humans perceive these problems as "dilemmas" rather than having clear-cut solutions.

Consequentialist mechanisms pose no theoretical difficulty because evolution favors adaptations based on the payoffs they produce. Kin selection, for example, favors mechanisms that maximize a weighted sum of individuals' payoffs based on relatedness (Hamilton, 1964). That is, Hamilton's theory is consistent with decision procedure (1), and observations in species such as the burying beetle support the theory. Similarly, reciprocity (Trivers, 1971), mutualism (Sachs, Mueller, Wilcox, & Bull, 2004), and costly signaling (Zahavi, 1975) can also favor consequentialist mechanisms with positive weights on others' payoffs.

In contrast, Kant's moral philosophy is described by decision procedure (2), and current theories do not straightforwardly explain what selection pressures give rise to it, leaving a gap in our understanding of human moral judgment. When people obey moral constraints, choosing actions other than a^*, as in (2), an explanation is required. What is the function of the constraining mechanism?

1.3. Does prohibiting beneficial acts generate benefits?

One common proposal is that simple moral rules of behavior such as "do not kill" or "do not sell sex" function to promote welfare (Gigerenzer, 2008). However, because these rules pertain to behavior *per se*, they necessarily prohibit beneficial actions in moral dilemmas when forbidden acts can yield net benefits. This raises the question: How can *prohibiting* beneficial acts *generate* benefits?

Psychologists have argued that Kantian rules of behavior, contrary to appearance, maximize welfare in the long run, on average, even if they lead to occasional errors (e.g., Gigerenzer, 2010). The idea is that calculating welfare consequences for specific cases is too computationally demanding, necessitating simple rules. This theory resembles the position in moral philosophy of "rule utilitarianism," in which a set of inflexible rules is observed because it is the best feasible way to maximize welfare (Sunstein, 2005).

In one version of this argument, moral constraints are implemented by emotions (Greene, 2007). The reluctance to kill in the context of moral dilemmas, on this view, is due to emotional systems that guide behavior (Haidt, 2001). That is, these emotions, whose function is to "motivate altruistic behavior" (Pyysiäinen and Hauser, 2010, p. 105), inhibit the choice of a^* when it is in W. In sum, *the predominant explanation for nonconsequentialism is that these judgments reflect the operation of human altruism systems that are implemented via moral rules of behavior.*

In contrast, we propose the alternative hypothesis that human altruism systems are consequentialist, as in (1) above, just like altruism mechanisms in burying beetles. If this is true, then nonconsequentialism in moral dilemmas is not due to the operation of altruism mechanisms. Instead, we have argued elsewhere that moral nonconsequentialism might be designed for strategic interactions among perpetrators, victims, and third-party condemners (DeScioli & Kurzban, 2009a). Here, however, we focus on the nature of human altruism systems, specifically whether or not these systems are consequentialist.

These two possibilities are shown in Fig. 1. The first possibility, depicted in the top panel, is that altruism

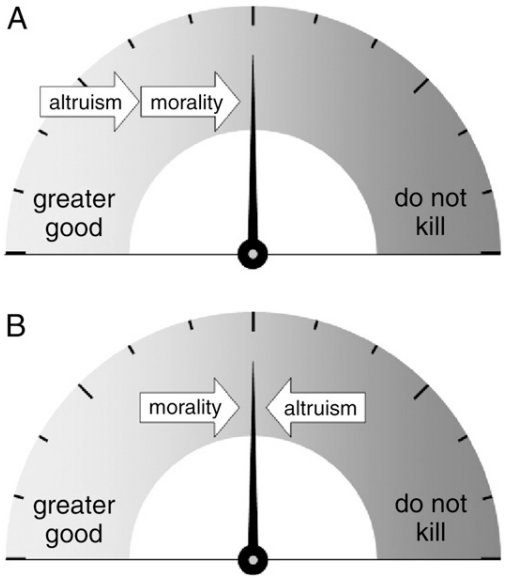

Fig. 1. Two models of decision-making in moral dilemmas. In moral dilemmas, there is a tension between maximizing inclusive fitness, the greater good, and adhering to moral rules such as "do not kill." In one model (panel A), evolutionary processes leading to altruism, such as kin selection or reciprocity, cause moral systems which implement simple heuristics for welfare, and these mechanisms push individuals to adhere to moral rules. In another model (panel B), moral cognition pushes behavior toward adherence to simple rules, but altruism systems counteract moral cognition, pulling behavior toward maximizing aggregate welfare.

mechanisms push people toward obeying moral constraints such as "do not kill" regardless of the consequences. An alternative view, depicted in the bottom panel of the figure, is that people have altruism systems that reflect the calculus of inclusive fitness (Burnstein, Crandall, & Kitayama, 1994) and therefore push people toward superior welfare outcomes; there are also moral systems, which serve a strategic function (DeScioli & Kurzban, 2009a), pushing in the opposite direction toward Kantian rules.

Consistent with the view that human altruism is consequentialist, Burnstein et al. (1994) found that people's tradeoffs among kin maximized inclusive fitness: People prefer to help larger numbers and more closely related relatives. However, this previous study did not investigate moral dilemmas (i.e., when a^* is in W). Hence, previous research cannot discriminate between decision processes using maximization based on payoffs (1) versus moral constraints (2). This requires studying cases where the action that maximizes inclusive fitness is labeled morally wrong.

1.4. The present studies

To test these alternative theories about how altruism works in moral dilemmas, we needed to manipulate a factor that, on existing theories, should influence the extent to which people rely on moral constraints, as in (2). If reasoning using these constraints is designed for directing altruism toward relatives, allies, and exchange partners, then by making these individuals the targets in a moral dilemma, *more* nonconsequentialist decision making should be observed.

In the footbridge version of the Trolley Problem, a majority of people report that they would not kill one person to save five people (Hauser, 2006). But what if all of the people involved were the participant's siblings? That is, what if participants had to decide whether they would kill one sibling to save five other siblings? People are more altruistic toward siblings than strangers. Current theories (Gigerenzer, 2010; Hauser, 2006) hold that human altruism operates via moral constraints. Therefore, the heuristic model predicts that the moral constraint "do not kill" will be applied more frequently in moral dilemmas among kin than among strangers.

Hypothesis A. An increase in participants' altruistic dispositions toward individuals in a moral dilemma will make them *more* likely to make decisions based on the moral constraint "do not kill" rather than the consequences.

The Hamiltonian, consequentialist model makes a different prediction. This model holds that human altruism systems are implemented by consequentialist systems of type (1) above. This idea leads to the prediction that when the moral dilemma requires balancing the welfare of one's relatives or allies, people will be more likely to reason according to consequentialism and hence *less* likely to use moral constraints such as "do not kill." Therefore, this theory predicts that greater altruism for kin will cause less use of the moral constraint "do not kill" in moral dilemmas among kin.

Hypothesis B. An increase in participants' altruistic dispositions toward individuals in a moral dilemma will make them *less* likely to make decisions based on the moral constraint "do not kill" rather than the consequences.

We note two additional points. First, we point out that Hypothesis A also applies to theories in which emotions are responsible for moral constraints. The emotion view predicts that increasing the emotional content by changing from strangers to kin will lead to greater use of the moral constraint. Second, Hypotheses A and B apply to the difference between strangers and friends given that we expect people to direct altruism toward friends more than strangers.

2. Study 1

Here we report studies in which we presented human subjects with problems not unlike those faced by nonhuman animals such as burying beetles. In particular, we asked participants whether they would kill one person in order to save five other people, varying the relationship between the participant and the other people in the situation (kin, friends, strangers). Note that our main dependent measures were people's judgments about their own behavior, or *conscience* (what they would do, whether their own actions would be wrong, etc.), rather than measures of participants' judgments

of others' behavior, or *condemnation*. The rationale for this approach is that the heuristic theories that we are testing are theories of how people make their own decisions (using welfare-maximizing heuristics) rather than how they will judge others' decisions, which is a very different task from a strategic perspective (DeScioli & Kurzban, 2009a). This experimental design allowed us to examine the relationship between cognitive systems involved in altruism with cognitive systems involved in moral judgment.

2.1. Subjects

Participants were recruited from Amazon's "crowdsourcing" Web site, mturk.com (see Buhrmeister, Kwang, & Gosling, 2011), which we have previously found to generate results similar to both college samples and a sample from a park in a large urban area in a study of social relationships (DeScioli & Kurzban, 2009b; see also Kurzban, Dukes, & Weeden, 2010). Participation was voluntary, compensated with a small amount of money—US$.15 for all the studies reported here—and anyone with access to the Web site could participate. We collected N=327 subjects roughly evenly divided among conditions (55% female; Table 1).

2.2. Methods

Our first interest was in investigating people's decisions across vignettes in which subjects faced the Trolley Problem when they could (a) kill one brother to save five brothers, (b) kill one friend to save five friends, and (c) kill one stranger to save five strangers. So, in a between-subject design, subjects were presented with the footbridge version of the Trolley Problem in which all of the people involved were strangers, friends of the subject, or brothers of the subject.

> Imagine that one day you are all walking near some trolley tracks. You are on a footbridge over the tracks. One person walks over and stands next to you. He is wearing a large, heavy backpack. Suddenly, a trolley is quickly approaching. You see that five other people are standing on the tracks. The only way to save them is to push a heavy object in front of the trolley. The only available heavy object is the man with the backpack (you are not heavy enough). There is not enough time to take off the backpack, and the people on the track are too far away to hear if you yell a warning. So, you have only two choices. If you push the man onto the tracks, then the trolley will be slowed and the five other people will be unharmed. You are forced to decide whether to push the man in front of the trolley, killing him, or to do nothing, allowing the five other people to die.

After reading the dilemma, subjects were asked a series of questions. Each block of questions had to be completed before subjects could move on to the next set of questions. Our main dependent measure was first, and asked subjects to report whether or not they would push and kill the person. The second two questions in the first set asked subjects to indicate both whether pushing the man onto the tracks was morally wrong *and also* whether not pushing the man onto the tracks was morally wrong. We asked both of these questions because we thought it possible that some participants would think that both pushing and not pushing were wrong.

The second set of questions asked subjects to think quantitatively about the tradeoff in the vignette. Specifically, we asked them to indicate how many people would have to be on the tracks for pushing to be morally permissible, as well as how few people would have to be on the tracks for *not* pushing to be morally permissible.

A third question in this set was simply to test participants' understanding, and asked whether pushing or not pushing would result in greater harm. (We did not analyze these responses.)

In the last set of questions, we asked a forced-choice item, asking participants to compare pushing and not pushing and asking which of the two was more morally wrong. Two additional questions asked participants to evaluate the moral wrongness of each act on a one to seven scale (as opposed to simply whether the act was wrong or not, as in the first set of questions). Finally, we asked what the subjects would want someone else to do if someone else instead of the subjects themselves were on the footbridge.

We ran two additional studies. The first was identical except that instead of the footbridge dilemma, we used the switch version of the Trolley Problem, in which the actor can pull a switch that will divert the trolley onto a side track, killing one person to save five. Second, we ran the original footbridge version, varying the relationships of the people in

Table 1
Descriptive statistics, Study 1, percentages and means (S.D.)

	Study 1: footbridge			Study 1: switch		
	Stranger	Friend	Brothers	Stranger	Friend	Brothers
N	111	100	116	96	96	97
Would you push/switch?	27.9%	41.4%	47.4%	77.1%	88.5%	88.7%
Is it wrong to push/switch?	85.6%	88.0%	84.5%	45.8%	40.6%	38.1%
Is it wrong not to push/switch?	60.4%	62.0%	64.7%	77.1%	67.7%	81.4%
How wrong is pushing/switching?	5.6 (1.8)	5.5 (1.9)	5.4 (1.8)	3.8 (2.1)	3.5 (2.0)	3.5 (2.1)
How wrong is not pushing/switching?	3.7 (2.3)	3.8 (2.3)	4.0 (2.3)	5.0 (2.0)	4.9 (2.0)	5.0 (2.2)
Is pushing/switching worse?	67.6%	59.0%	62.9%	27.1%	24.0%	17.5%
Would you want someone else to push/switch?	44.1%	48.0%	61.2%	81.2%	89.6%	89.7%

Note: For binary questions, the percentage responding "yes" is reported.

the vignette to the subject, including the case that we predicted would lead to the most consequentialist responding, when one can push one stranger to save five siblings.

After completing the questions, subjects were asked some demographic questions, including age, sex, location, race, religion, and number of siblings.

2.3. Results

2.3.1. Footbridge dilemma

Summary statistics for the primary dependent measures of interest are presented in Table 1.

2.3.1.1. What would you do? We first looked at the primary question of interest: did the relationship to the subjects of the people named in the vignette—strangers, friends, or brothers—affect subjects' reports of whether they would push the person with the backpack off of the footbridge? Because this was a binary dependent variable, we conducted a logistic regression. We entered three predictors into the model: stranger vs. brother/friend, brothers vs. friend, and sex (Table 2). [In this and all subsequent models, we did not include siblings as a predictor because few (11.9%) participants reported having no siblings, undermining the utility of this variable as a covariate.] The odds of pushing increased significantly when the scenario involved brothers or friends, rather than strangers (Table 2). There was no significant difference between the friend and brothers conditions. There was no effect of sex.

2.3.1.2. Is (not) pushing wrong? In contrast to how people said they would act, their moral judgments of pushing did not show differences across conditions. We used logistic regression to test for differences in moral judgments of pushing, *not pushing*, and the forced-choice item about which is worse. We entered three predictors into each model: stranger vs. other, brothers vs. friend, and sex.

Neither the model predicting whether pushing was wrong (model χ^2_3=2.05, p=.56) nor the model predicting whether pushing was not wrong (model χ^2_3=4.69, p=.20) provided an adequate fit to the data. The same was the case for a logistic regression in the forced-choice version of this question in which subjects had to indicate which was more morally wrong, pushing or not pushing (model χ^2_3=4.36, p=.23).

In order to investigate how morally wrong subjects thought pushing was on the seven-point scale, we conducted a 3×2 analysis of variance (ANOVA) with relationship (stranger, friend, brothers) and sex as the independent variables. There was no main effect of type of relationship ($F_{2,319}$=.15, p=.86), but there was a main effect of sex ($F_{1,319}$=11.52, p=.001). Women thought that pushing was more morally wrong (M=5.78) than did men (M=5.10). The interaction of type of other and sex was not significant ($F_{2,319}$=1.87, p=.16). We also conducted a 3 (relationship)×2 (sex) ANOVA predicting how morally wrong it was not to push. There were no main effects of type of relationship ($F_{2,319}$=.79, p=.45) or sex ($F_{1,319}$=1.11, p=.29), nor was there a significant interaction ($F_{2,319}$=1.73, p=.18).

2.3.1.3. Quantitative Items. Data regarding how many people would have to be on the tracks for pushing to be permissible ranged extremely widely, from zero to infinite. A nontrivial proportion of subjects gave the peculiar response of zero, indicating that it would be morally acceptable to push even if no one were on the track (11%, 19%, and 8% for brothers, friends, and strangers, respectively). A small number of subjects gave a response of one (3%, 1%, and 5%). Some subjects gave the consequentialist response of 2 (22%, 20%, and 19% for brothers, friends, and strangers, respectively). But the bulk of subjects responded in more or less Kantian fashion, with large numbers, including explicit responses such as "no amount" or a string of 9's or 1's followed by many zeros. The number of subjects indicating a value of three or more was 59%, 60%, and 65% for brothers, friends, and strangers, respectively.

2.3.1.4. What would you want others to do? Finally, we conducted a logistic regression on the question asking what the subject would want someone else to do in this situation (model χ^2_3=9.36, p=.025). There were significant differences, such that people indicated a greater desire that the person on the footbridge be pushed when the people in the vignette were friends or brothers compared to the case in which they were strangers. We similarly found a difference on this measure for brothers versus friends, in the expected direction, with people reporting that they would want others to push in the case of brothers more than in the case of friends (Table 2).

2.3.1.5. Choosing wrongful behavior. In the cases of brothers and friends, one can deliver aggregate benefits to kin and allies by killing one to save five. So, if altruism systems can counter moral systems, then willingness to push *despite perceiving the act as wrong* should increase for brothers and friends relative to strangers. Indeed, the fraction

Table 2
Logistic regression, Study 1, footbridge dilemma

Variable	β [a]	S.E.	Wald χ^2	p	Exp(β)
Would push					
Stranger vs. brothers/friend	0.85	0.29	8.85	.003	2.32
Brothers vs. friend	−0.27	0.28	0.94	.334	0.76
Sex	0.38	0.23	2.60	.11	1.46
Want other to push					
Stranger vs. brothers/friend	0.74	0.27	7.39	.007	2.10
Brothers vs. friend	−0.60	0.28	4.60	.031	0.54
Sex	0.18	0.23	0.59	.44	1.46

Note. Effect tests for logistic model. Model fit for "Would push": $\chi^2(3)$ =12.8, p=.005. Model fit for "Other push": $\chi^2(3)$=9.4, p=.025.

[a] Logistic regression coefficient. The exponential of β is the change in the odds of pushing the person with the backpack off of the footbridge.

of subjects who reported *both* that they would push *and* that pushing is morally wrong was indeed greater for brothers (36%) than strangers (17%), χ^2_1=11.6, p<.001, and was also greater for friends (32%) than strangers, χ^2_1=7.1, p<.01.

Furthermore, many participants reported *both* that it is wrong to push *and* that it is wrong to not push, leaving them with no morally sanctioned options. We observed that both alternatives were judged wrong by 50% for strangers, 54% for friends, and 57% for brothers. This finding cautions that the wrongness of one alternative does not imply that participants view the other alternative as morally sound.

2.3.2. Switch version

Procedures were identical for the switch version of the dilemma, except that the vignette was changed. In the switch version, the five people on the track can be saved by throwing a switch that diverts the trolley to a sidetrack, but there is a person standing on the sidetrack who will be killed. As in the footbridge version, the person on the track has a large backpack, which is sufficiently heavy to stop the trolley. Participants were recruited from the same Web site and again paid US$.15 for their participation.

We used the same between-subjects design as in the first study, and subjects could participate in only one condition. We collected data from N=289 subjects (55% female) roughly evenly divided among conditions.

2.3.2.1. What would you do? Did the subjects' relationship to the people named in the vignette—strangers, friends, or brothers—affect subjects' reports of whether they would pull the switch? Again, we conducted a logistic regression using the same predictors as above (Table 3); the odds of pulling the switch compared to not pulling the switch increased significantly when the scenario involved brothers or friends, mirroring the effect found in the footbridge dilemma in direction and magnitude. There was no significant difference between the friend and brothers conditions. There was also a main effect of sex: female subjects were less likely to say that they would pull the switch.

2.3.2.2. Is (not) switching wrong? We conducted a similar analysis for the item asking if pulling the switch was wrong using the same model as above. This model (model χ^2_3=12.4, p=.006) revealed only one significant effect: a large sex difference [β=−.82, S.E.=.25, Wald χ^2_1=10.6, p=.001, Exp(β)=.44], showing that female subjects judged pulling the switch to be more wrong than male subjects did. The model predicting whether pulling the switch was not wrong did not provide a good fit (model χ^2_3=6.12, p=.11).

Looking at how wrong subjects viewed pulling the switch, we conducted a 3×2 ANOVA, as above. There was no main effect of type of relationship ($F_{2,275}$=.94, p=.39), but there was again a main effect of sex ($F_{1,275}$=11.49, p=.001). Female subjects thought that pushing was more morally wrong (M=3.95) than male subjects did (M=3.12). The interaction of relationship and sex was not significant ($F_{2,275}$=0.80, p=.45).

We also conducted a 3 (relationship) × 2 (sex) ANOVA predicting how morally wrong it was *not* to pull the switch. As in the footbridge version, there were no main effects of relationship, ($F_{2,274}$=.10, p=.91) or sex ($F_{1,274}$=0.18, p=.67), nor was there a significant interaction ($F_{2,274}$=1.42, p=.25).

As in the footbridge version, a logistic regression in the forced-choice question, in which subjects had to indicate which was more morally wrong, switching or not switching, revealed a nonsignificant model fit (χ^2_3=2.61, p=.46).

2.3.2.3. What would you want others to do? Finally, we conducted a logistic regression on the question asking what the subject would want someone else to do in this situation. The fit of the model was poor (Table 3).

2.3.2.4. Choosing wrongful behavior. A substantial fraction of subjects indicated that they would push and that pushing was wrong, though these fractions were nearly identical across conditions (29%, 31%, and 30% for brothers, friends, and strangers, respectively).

Table 3
Logistic regression, Study 1, switch version

Variable	β [a]	S.E.	Wald χ^2	p	Exp(β)
Would switch					
Stranger vs. brothers/friend	0.96	0.41	5.34	.02	2.61
Brothers vs. friend	−0.13	0.46	0.08	.78	0.88
Sex	0.72	0.36	4.05	.04	2.05
Want other to switch					
Stranger vs. brothers/friend	0.74	0.44	2.86	.09	2.11
Brothers vs. friend	0.09	0.51	0.03	.86	1.09
Sex	0.04	0.38	0.01	.92	1.04

Note. Effect tests for logistic model. Model fit for "Would switch": $\chi^2(3)$ =10.6, p=.014. Model fit for "Other switch": $\chi^2(3)$=4.6, p=.21.

[a] Logistic regression coefficient. The exponential of β is the change in the odds of pushing the person with the backpack off of the footbridge.

Table 4
Descriptive statistics, Study 1, mixed relationships, percentages and means (S.D.)

Push to save	Stranger Friends	Friend Brothers	Stranger Brothers
N	106	120	94
Would you push?	52.8%	54.2%	56.4%
Is it wrong to push?	84.9%	84.2%	85.1%
Is it wrong not to push?	67.0%	72.5%	66.0%
How wrong is pushing?	5.5 (1.7)	5.3 (2.0)	5.8 (1.6)
How wrong is not pushing?	4.0 (2.3)	4.2 (2.4)	3.9 (2.3)
Is pushing worse?	61.3%	56.7%	63.8%
Would you want someone else to push?	67.0%	60.8%	73.4%

Note: For binary questions, the percentage responding "yes" is reported.

2.3.3. Footbridge dilemma, mixed relationships

We ran a follow-up study investigating what people would do if they were faced with a decision to push (1) one stranger to save five friends, (2) one stranger to save five brothers, and (3) one friend to save five brothers.

Procedures were otherwise identical, and we collected data from 320 subjects (57% female) (Table 4).

All of the models showed a poor fit to the data with one exception. How wrong it was not to push showed a significant model fit (χ^2_3=8.55, p=.036), but the only significant effect was the sex difference: women are less likely than men to say that it is wrong not to push [β=−.66, Wald χ^2_1=7.06, p=.008, Exp(β)=.52].

2.3.3.1. Choosing wrongful behavior. Again, many subjects indicated that they would push despite reporting that they thought pushing was morally wrong; these fractions were very similar across conditions, however (40%, 43%, and 39% for stranger/friend, stranger/brothers, and friend/brothers, respectively).

2.4. Discussion

The results of Study 1 show that the subjects' relationship to the person on the footbridge affects their decisions about whether to push and kill the person. People are more likely to report that they would push one brother/friend off of the footbridge to save five brothers/friends, and they similarly report an increased desire that others do so.

Moral judgments, however, do not change. It is as morally wrong to push a stranger off of a footbridge to save five strangers as it is to push one brother off of a footbridge to save five brothers.

In short, when brothers or friends are involved in the footbridge dilemma, people report a greater willingness to do the morally wrong thing: Hamilton makes people less Kantian. These findings contradict the prediction, derived from heuristic models, that increasing altruistic dispositions will increase adherence to Kantian rules. On the other hand, the results support models in which altruism systems are distinct from, and sometimes oppose, cognitive mechanisms underlying rule-based moral cognition (DeScioli & Kurzban, 2009a).

We replicated this effect in the switch version even though for the key dependent variable, willingness to switch, results were near ceiling, as in similar studies (Hauser, 2006). What subjects reported they would do varied with the relationship of the people involved in the vignette; judgments of wrongness showed no such effect.

Finally, results in vignettes that mixed the relationship to the subject were surprising in that even when subjects were asked if they would push one stranger to save five brothers, results were distant from the ceiling, with 44% of people still unwilling to push. This result illustrates the strength of "Kantian" psychology and the degree to which it undermines kin selected systems. It is worth considering these results in the context of Haldane's quip that he would sacrifice himself for two brothers. Our results imply the biologically odd possibility that people (report that they) are willing to die, but not kill, in order to save multiple siblings.

3. Study 2

Trolley Problems are appealing because of the wealth of data gathered and because their features can be varied to address the details of the vignette that carry moral weight, affecting decisions (e.g., Mikhail, 2007). In Study 2, we conducted a second set of studies to test whether the result replicates and is found in additional content domains. First, we replicated the footbridge dilemma exactly as in the previous study, but with two people on the tracks instead of five. Second, we used a vignette similar to the footbridge dilemma, but with a different source of peril and a different means of rescue (see below). Finally, we wanted to see if the effect only occurred when the tradeoff was a matter of life and death, and so we used another scenario to investigate the effect when the people involved in the vignette are threatened with nonfatal injury. Because friends and brothers behaved similarly in Study 1, for simplicity, we reduced the independent variable to simply strangers and brothers in Study 2.

3.1. Subjects

Participants were again recruited from the same commercial Web site (Amazon's mturk) as in Study 1 and were paid US$.15 for their participation.

3.2. Methods and results

3.2.1. Footbridge dilemma, two brothers

We modified the vignette in the first study, placing two people on the track instead of five. Procedures were otherwise identical, and we collected data from 108 subjects (45 male) evenly divided between the two conditions. For this and all subsequent experiments, we first ran regressions looking at condition, sex, and their interaction. There were no significant interactions between condition and sex on any variable, and so below, for simplicity, we report χ^2 and t tests. Note that there were two sex effects: female subjects were less likely to say that they would push (19% vs. 38%, χ^2_1=4.7, p<.05), and they said that pushing was more wrong (5.8 vs. 5.0, t_{106}=2.4, p<.05). There were no other effects of sex in this experiment or the other two experiments in this section.

Replicating our primary result, a larger proportion of people reported that they would push the one person to save two in the brother condition (39%) than in the stranger condition (15%), χ^2_1=12.6, p<.001.

On the seven-point scale, pushing one brother to save two was seen as less wrong (5.1) than pushing one stranger to save two (5.9), t_{106}=−2.49, p<.01. Not pushing a brother (3.8) was seen as no more wrong than not pushing a stranger

(3.3), t_{106}=1.23, n.s. Looking at the binary measures, there was no difference in the fraction of subjects reporting that it was wrong to push in the brother case (83%) compared to the stranger case (87%), and similarly for not pushing in the brother case (61%) compared to the stranger case (44%), both p's>.05.

More people in the brother case said that they would want someone else to push (57%) than in the stranger case (22%), χ^2_1=13.3, p<.001. In addition, a greater fraction of subjects indicated that they would push and that pushing was morally wrong (24%) than in the stranger case (7%; χ^2_1=5.6, p<.05).

In sum, we replicate the primary effects of interest and find that more people report that they would push and report wanting someone else to push in the brother case. Wrongness judgments were largely replicated, with an exception such that pushing in the brother case was seen as less wrong on the seven-point scale.

3.2.2. Counterweight dilemma

In the counterweight dilemma (see Appendix available on the journal's website at www.ehbonline.org), the reader is again on a footbridge and can only save two people in a ravine by pushing a man with a backpack off the bridge. In this vignette, the man is attached by rope to two men below about to be killed by a flash flood; by pushing the man, he is used as a counterweight, raising the two men in the ravine to the footbridge, leading to his death.

Procedures were identical, and we collected data from 140 subjects (61 male) roughly evenly divided between the two conditions (68 in the brother condition, 72 in the stranger condition). (Note that not all subjects answered all questions, leaving a small amount of missing data; this is reflected in the small differences in degrees of freedom.)

Once again replicating the primary effect of interest, a larger proportion of people reported that they would push the one person to save two in the brother condition (27/68, 40%) than the stranger condition (15/70, 21%), χ^2_1=5.44, p<.05.

On the seven-point scale, pushing one brother to save two (5.58) was seen as equally wrong as pushing one stranger to save two (5.43), t_{132}=−0.53, n.s. Not pushing a brother (3.78) was seen as equally wrong as not pushing a stranger (3.13), t_{132}=−1.88, n.s. Looking at the binary measures, there was no difference in the fraction of subjects reporting that it was wrong to push in the brother case (56/68, 82%) compared to the stranger case (62/72, 86%), and similarly for not pushing in the brother case (39/68, 57%) compared to the stranger case (31/72, 43%), both p's>.05.

More people in the brother case said that they would want someone else to push (35/68, 51%) than in the stranger case (24/69, 35%), χ^2_1=3.89, p<.05. However, the fraction of subjects reporting that they would push despite the fact that pushing was morally wrong did not differ between the stranger and brother conditions (28% and 17%, respectively).

In sum, while the wrongness of pushing and not pushing did not differ between brothers and strangers, people indicated that they themselves were more likely to push and that they would want someone else to push when the people involved were brothers compared to strangers.

3.2.3. Windstorm dilemma

In the windstorm dilemma (see Appendix available on the journal's website at www.ehbonline.org), one person is climbing down from a tree in a sudden, dangerous windstorm. Two other people are in jeopardy and can only be made safe by shaking the person off the ladder and bringing it to the other location. This dilemma explores if the same effect occurs when the situation is not a case of life and death. Procedures were identical, and we collected data from 106 subjects (41 male) from the same online Web site (mturk), restricted to subjects in the United States,[1] roughly evenly divided between the two conditions (56 in the brother condition, 50 in the stranger condition).

No significant differences in any of the dependent measure emerged. Results were in fact very similar across treatments. In the strangers condition, 28% indicated that they would shake the ladder, compared to 32% in the brothers condition. In the strangers condition 76% indicated that shaking the ladder was wrong; this value was 71% in the brothers condition. Wrongness judgments on the seven-point scale for moving the ladder were nearly identical in the strangers condition (M=5.0) and the brothers condition (M=5.0). These judgments for the wrongness of failing to move the ladder were also similar (strangers, M=3.7; brothers, M=3.8). Nearly the same fraction of subjects indicated that they would want someone else to move the ladder in the strangers condition (42%) as in the brothers condition (41%). The fraction of subjects indicating that they would shake the ladder and reporting that doing so was morally wrong did not differ between conditions (11% and 24% in the brother and stranger treatments, respectively). None of the relevant tests approached significance (all p's>.05).

The increased tendency to choose to perform the harm-reducing option when brothers were involved did not occur in the context of nonfatal harm.

3.3. Discussion

The effect observed in the first study was replicated when the number of people on the trolley tracks was reduced from five to two and when the vignette was changed in superficial ways. That is, people were more likely to report that they would kill one to save more than one when brothers were involved. However, there was no change in moral judgments

[1] Preliminary data indicated a strong cross-cultural difference in this vignette, so we conducted this study with a US-only sample. People in India (N=46) were much more likely to say that they would shake the stranger off the ladder (67%) than the US sample (28%, N=50), χ^2_1 = 14.9, p<.001. There were too few Indians to conduct a similar analysis for the other studies reported here. Removing these observations from the other studies reported here does not substantively alter the findings.

across treatments. This effect did not occur in the context of a similar vignette in which victims were harmed but not killed.

4. General discussion

4.1. Summary of results

The studies presented here show that an individual's reported willingness to cause the death of one person to save many people depends on the individual's relationships with the people in the situation. When friends and relatives are involved, subjects are more consequentialist—ignoring the Kantian prohibition against causing death. Participants report that they are more willing to kill one brother to save five brothers, or one friend to save five friends, than they are to kill one stranger to save five strangers. Importantly, however, the subject's relationships did not alter judgments of moral wrongness. Subjects judged pushing one brother to save five brothers as no more or less morally wrong than pushing one stranger to save five strangers.

Recall the theory that heuristics for altruism are responsible for nonconsequentialist choices in moral dilemmas (Greene, 2007; Sunstein, 2005). This theory predicts that increasing participants' altruistic disposition toward people in trolley dilemmas should increase use of the heuristic "do not kill." However, we find the reverse. Participants are less likely to adhere to the moral rule "do not kill" in decisions about kin and friends than in decisions about strangers. These findings provide evidence against models in which rule-based moral judgment is caused by altruism systems shaped by kin selection or reciprocal altruism.

Instead, our results support models in which altruism systems differ from, and sometimes oppose, moral systems (DeScioli & Kurzban, 2009a). In this view, humans have cognitive systems for altruism shaped by kin selection and reciprocity, and, as in other species, these systems perform nuanced computations aimed at maximizing inclusive fitness. However, humans also have cognitive systems for moral judgment, and it is these moral mechanisms that focus on inflexible rules of behavior rather than welfare outcomes. In dilemmas such as the Trolley Problem, altruism systems and moral systems oppose one another, with altruism systems pulling toward welfare maximization and moral systems pushing toward adherence to Kantian rules.

These conclusions are also supported by participants' wrongness judgments. Intriguingly, while participants' decisions about their actions were influenced by relationships, their moral judgments were not. This suggests that, at least under some conditions, mechanisms computing wrongness discard—or at least are unaffected by—the individual's relationships. This implies that moral judgment has some degree of *impartiality* (DeScioli & Kurzban, 2009a; Lieberman & Linke, 2007).

Another interesting result from the present studies is that people frequently report that both options in a dilemma are immoral. This observation strongly suggests that in making moral judgments, people do not evaluate the possible candidate acts against one another and perform computations about their relative position on some dimension. For example, it could have been that the least harmful act was judged moral or, perhaps, drawing on intuitionist models, that the least emotionally laden option was judged to be not immoral. This is clearly not the case. This feature of moral cognition highlights how ill-equipped it seems for handling the types of welfare tradeoffs routinely faced by organisms such as burying beetles: The output that all options are morally wrong seems poorly designed for guiding an organism's behavior toward beneficial outcomes.

A surprising result is that nearly half of our subjects reported that they would be unwilling to push one stranger to save five brothers. This might of course reflect some type of self-report bias, but if we take these reports at face value, it implies that the mechanisms that inhibit immoral action act as an extremely powerful counterweight to kin selected systems. To put this in comparative perspective, if it were found that mother bears, for instance, routinely preferred five of their cubs to be killed rather than kill an *unrelated* bear, this would be considered a biological oddity of the first order, and kin selection would be very obviously unable to explain this result. Indeed, in numerous species, spanning taxa—insects, birds, fish, and mammals—individuals directly cause or allow the deaths of relatives to benefit one or more other relatives (Mock, 2004). Why are humans reporting that they are unwilling to act similarly?

4.2. Implications for theories of moral judgment

Why do humans look so different from nonhumans in the context of these kinds of choices? It seems unlikely that human minds are so much less sophisticated than insect minds that they must implement altruism using simple heuristics such as "do not kill." Further, even if one were to grant this possibility, the heuristics previously proposed cannot explain the pattern of results. Heuristics such as "do not tamper with nature" (Sunstein, 2005, p. 540) are too vague, and other candidate heuristics do not successfully predict judgments across moral dilemmas. More generally, it seems likely that human moral cognition is highly complex and sophisticated, casting doubt on theories that posit simple mechanisms. Mikhail (2007), for instance, has shown how the personal/impersonal distinction (Greene & Haidt, 2002; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001) does not account for the observed patterns of data. Successful theories will have to account for the subtle nuance and texture of moral judgment (Mikhail, 2005). The data reported here add to these difficulties, showing a pattern of data that is the opposite of what current heuristic models predict.

We conclude that while kin selected psychology has been considered to be a kind of morality, we think that it is more productive to think of kin selected systems in terms of their particular function—delivering benefits to close genetic relatives—and that there are, in addition, moral systems, and

these systems have a distinct function or functions. The fact that it is possible to set moral mechanisms and kin selected systems against one another implies that the idea that one is a subcategory of the other is incorrect.

4.3. Alternative interpretation

An alternative interpretation (suggested to us by an anonymous reviewer) is that indirect reciprocity (Nowak & Sigmund, 2005) explains why people report greater willingness to push brothers than strangers in dilemmas such as the footbridge problem. Indirect reciprocity involves individuals who track others' reputations and seek to maintain their own positive reputation for cooperation. If pushing a brother were less damaging to one's reputation as a cooperator than pushing a stranger, then the stranger–brother difference could be driven by reputational concerns. Might pushing one stranger to save five be seen as more *selfish* than pushing one brother to save five?

One reason to doubt this interpretation is that wrongness judgments from the present experiments show that people did not view pushing a stranger as more *wrong* than pushing a brother. This lack of a difference suggests that people are viewed as equally morally culpable for pushing strangers and brothers in these situations. If the brother–stranger difference is to be explained by reputation, it seems not to be driven by individuals' *moral* reputations.

Further, consider that the success of cooperators in indirect reciprocity models is due to the surplus of benefits generated by cooperative acts. In the footbridge dilemma, pushing generates surplus benefits, which seems to commit these models to regarding pushing (net effect=5−1=+4 lives) as "cooperation" and refusal to push (net effect=1−5=−4 lives) as "defection." A person seeking a reputation as a cooperator, then, should prefer to be seen as a person who pushes rather than a person who refuses to push. If reputation is driving these decisions, then refusing to push does not produce a reputation as a cooperator who promotes welfare, but rather, as someone who complies with moral constraints *even when doing so destroys welfare.*

Crucially, indirect reciprocity models turn on realized benefits to the targets of behavior, independent of the relationship between the actor and these targets. Hence, this logic applies whether the targets of the behavior are kin or strangers. If decisions to push were driven by concern for reputation as an altruist, then there should be no difference in pushing decisions between the two conditions.

So, because pushing—strangers or brothers—is cooperative in the sense required by indirect reciprocity, models of this type predict that people will seek a reputation as one who pushes rather than one who refuses to push. Consequently, indirect reciprocity faces a stiff challenge to explain refusal to push; populations of agents who do not push—and reward nonpushers—are unstable against pushers because of the net benefits of pushing. For an indirect reciprocity model to explain choices in the footbridge dilemma, an explanation would be required for how the refuse-to-push strategy could overcome the large numerical advantages of the pushing strategy (Maynard Smith, 1982).

More generally, tradeoffs among others' welfare are widespread in animals ranging from burying beetles to pelicans to humans (e.g., Burnstein et al., 1994; Mock, 2004). Individuals who were able to make these tradeoffs would have had a considerable advantage, and indeed, many animals have evolved mechanisms for making nuanced tradeoffs in these situations. To the extent that humans depart from Hamiltonian predictions, an explanation is required in terms of some other adaptive problem that human mechanisms are designed to solve. Reciprocity models are unsatisfying because they are based on reaping gains in trade, which is precisely what human decisions fail to accomplish in these problems. Better explanations might be found by considering the many other strategic games, beyond capturing gains in trade, that arise in human social life (e.g., Schelling, 1960).

A final reason to question an indirect reciprocity explanation is the difference between footbridge and switch problems. In both the switch and footbridge problems, the decision is whether to cause one death to save five. The welfare consequences of pushing and switching are equivalent. However, people report greater willingness to kill one person to save five people in the switch problem, suggesting that something other than reputation *as cooperators* (i.e., producers of benefits) is driving decisions.

In sum, while we are sympathetic to the general notion that decisions in these vignettes might be driven in part by computations surrounding reputation, indirect reciprocity models do not seem easily able to account for the patterns of data we observe here.

4.4. Conclusion

In short, we conclude that moral cognitive systems, far from being designed to generate aggregate benefits, act to substantially reduce aggregate welfare in the tradeoff situations investigated here. So, if altruism theories of morality are incorrect, then moral judgment is caused by a different cognitive system with a different (nonaltruism) function.

What, then, is the function of the computational system that delivers Kantian, nonconsequentialist moral judgments? DeScioli and Kurzban (2009a) suggested that the function of this system might be comprehensible in the context of the multiplayer strategic dynamics that occur among third-party condemners, perpetrators, and victims in interactions involving moral violations. Specifically, an important adaptive problem for third parties is to coordinate their condemnation decisions with others (DeScioli, Bruening, & Kurzban, 2011). If an individual condemns a wrongdoer alone, then there is a greater risk of retaliation than if many third parties condemn the wrongdoer.

According to this view, condemnation is not designed to reduce welfare losses, but rather to avoid discoordination.

Suppose that an agent's goal is to condemn those acts and only those acts that other agents similarly condemn. In such a case, agents should use whatever structural features of the moral situation that others are using, *independent of the welfare consequences of those structural features*. In the context of our notation above, moral rules specify action constraints *W*, which function as coordination points for condemnation. Moral intuitions, on this view, can be very complex (e.g., Mikhail, 2007) as long as they are shared because, just like symbols, they get their value from consensus. This view suggests that moral dilemmas [when (1) and (2) are in conflict, above] can be understood as the tension between altruism systems and (Kantian) coordination systems (Fig. 1).

This line of reasoning implies that when people themselves make decisions with potential moral weight, they should seek to avoid actions in the set *W* to decrease the risk of a coordinated moral attack against themselves. These decisions must, of course, be set against other relevant concerns, such as those associated with kin altruism, as was the case here. Decisions weigh these factors against one another, which explains why choices over kin pull against the Kantian option, increasing the chance of choosing the action in set *W*, in this case, pushing one to save five in the footbridge dilemma.

Broadly, the strategic framework surrounding coordination focuses attention on third-party condemnation as a potential explanation for people's adherence to moral constraints. This approach, in turn, might illuminate why in the context of moral dilemmas, humans, but not burying beetles, appear Kantian rather than Hamiltonian.

Supplementary Materials

Supplementary materials related to this article can be found online at doi:10.1016/j.evolhumbehav.2011.11.002.

References

Buhrmeister, M. D., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*, 3–5.

Burnstein, E., Crandall, C., & Kitayama, S. (1994). Some neo-Darwinian decision rules for altruism: weighting cues for inclusive fitness as a function of the biological importance of the decision. *Journal of Personality and Social Psychology*, *67*, 773–789.

Cosmides, L., & Tooby, J. (1994). Beyond intuition and instinct blindness: the case for an evolutionarily rigorous cognitive science. *Cognition*, *50*, 41–77.

DeScioli, P., Bruening, R., & Kurzban, R. (2011). The omission effect in moral cognition: toward a functional explanation. *Evolution and Human Behavior*, *32*, 204–215.

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

DeScioli, P., & Kurzban, R. (2009b). The alliance hypothesis for human friendship. *PLoS ONE*, *4*, e5802.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, *5*, 5–15.

Gigerenzer, G. (2008). Rationality for mortal: how people cope with uncertainty. New York, NY: Oxford University Press.

Gigerenzer, G. (2010). Moral satisficing: rethinking moral behavior as bounded rationality. *Topics in Cognitive Science*, *2*, 528–594.

Greene, J. D. (2007). The secret joke of Kant's soul. In W. Sinnott-Armstrong, Ed. *Moral psychology, vol. 3: the neuroscience of morality: emotion, disease, and development* (pp. 35–79). Cambridge, MA: MIT Press.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. (2001). An fMRI investigation of emotional engagement in moral Judgment. *Science*, *293*, 2105–2108.

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, *6*, 517–523.

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

Hamilton, W. D. (1964). The genetic evolution of social behavior. *Journal of Theoretical Biology*, *7*, 1–52.

Hauser, M. D. (2006). Moral minds: how nature designed our sense of right and wrong. New York, NY: Ecco/Harper Collins.

Hauser, M., Young, L., & Cushman, F. (2008). Reviving Rawls' linguistic analogy. In W. Sinnott-Armstrong, Ed. *The cognitive science of morality: intuition and diversity* (pp. 107–155). Oxford, England: Oxford University Press.

Kurzban, R., Dukes, A., & Weeden, J. (2010). Sex, drugs, and moral goals: reproductive strategies and views about recreational drugs. *Proceedings of the Royal Society-B*, *277*, 3501–3508.

Lieberman, D., & Linke, L. (2007). The effect of social category on third party punishment. *Evolutionary Psychology*, *5*, 289–305.

Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge: Cambridge University Press.

Mikhail, J. (2005). Moral heuristics or moral competence? Reflections on Sunstein. *Behavioral and Brain Sciences*, *28*, 557–558.

Mikhail, J. (2007). Universal moral grammar: theory, evidence and the future. *Trends in Cognitive Sciences*, *114*, 143–152.

Mock, D. W. (2004). More than kin and less than kind: the evolution of family conflict. Cambridge, MA: Oxford University Press.

Nowak, M., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*, 1291–1298.

Pyysiäinen, I., & Hauser, M. (2010). The origins of religion: evolved adaptation or by-product? *Trends in Cognitive Sciences*, *14*, 104–109.

Sachs, J. L., Mueller, U. G., Wilcox, T. P., & Bull, J. J. (2004). The evolution of cooperation. *The Quarterly Review of Biology*, *79*, 135–160.

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

Sunstein, R. C. (2005). Moral heuristics. *Behavioral and Brain Sciences*, *28*, 531–543.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, *46*, 35–57.

Zahavi, A. (1975). Mate selection—a selection for a handicap. *Journal of Theoretical Biology*, *53*, 205–214.