



PROJECT MUSE®

Hidden Paths from Morality to Cooperation: Moral Judgments Promote Trust and Trustworthiness

Brent Simpson, Ashley Harrell, Robb Willer

Social Forces, Volume 91, Number 4, June 2013, pp. 1529-1548 (Article)

Published by Oxford University Press



➔ For additional information about this article

<http://muse.jhu.edu/journals/sof/summary/v091/91.4.simpson.html>

Hidden Paths from Morality to Cooperation

Hidden Paths from Morality to Cooperation: Moral Judgments Promote Trust and Trustworthiness

Brent Simpson and Ashley Harrell, *University of South Carolina*
 Robb Willer, *Stanford University*

Classic sociological solutions to cooperation problems were rooted in the moral judgments group members make about one another's behaviors, but more recent research on prosocial behaviors has largely ignored this foundational work. Here, we extend theoretical accounts of the social effect of moral judgments. Where scholars have emphasized the roles of moral judgments in clarifying moral boundaries and punishing deviants, we present two less intuitive paths from moral judgments to social behavior. We argue that those who engage in moral judgments subsequently act more morally. Further, we argue that group members anticipate the more moral behavior of judges, trusting them more under situations of risk and uncertainty. We thus establish paths from moral judgments to the primary foundations of voluntary cooperation: *trust* and *trustworthiness*. The results of three experiments support the predicted effects: Participants randomly assigned to make moral judgments were more trustworthy in subsequent interactions (Study 1). A follow-up experiment sought to clarify the underlying mechanism, showing that making moral judgments led individuals to view themselves as more moral (Study 2). Finally, audience members anticipated the greater trustworthiness of moral judges (Study 3).

Conflicts between individual and collective interests are fundamental to the structure of human societies (Hechter 1987; Coleman 1990; Kollock 1998). From everyday interpersonal exchanges to the mobilization of social movements, situations often occur where what is best for the individual may not be the most beneficial course of action for groups. Understanding the mechanisms through which groups resolve these tensions and motivate cooperation and prosocial behavior is a critical issue for the behavioral sciences (Hardin 1982; Wrong 1994; Fukuyama 1995).

Classical sociological thinking addressed a variety of processes through which the moral judgments individuals make about fellow group members' behaviors affect cooperation and prosocial behavior. For instance, Durkheim (1893) and others (Tocqueville 1835; Erikson 1966) viewed the moral judgments of

This research was supported by grants SES-0647169 and SES-1058235 from the *National Science Foundation*.

community members as critical for aligning and reaffirming moral boundaries, thus fostering group solidarity and pro-community behaviors.

But contemporary approaches to the puzzle of prosocial behavior have largely neglected these insights into how moral judgments increase prosocial behavior. Dominant theoretical approaches, especially those from the biological and economic sciences, have instead sought to explain cooperation and other forms of prosocial behavior largely in terms of material and reputational self-interest. For instance, recent research on punishment systems (Fehr and Gächter 2002) shows how material sanctions can reduce the conflicts between individual and collective interests, leading would-be free riders to contribute to the provision of public goods. Similarly, research on reputations (Barclay and Willer 2007) typically views prosocial behavior as motivated by pursuit of the downstream benefits of a good reputation. These approaches predict that individuals will act egoistically when they can escape punishment for their selfishness (Hechter 1987) and when prosociality does not yield reputational benefits (Semmann, Krambeck and Milinski 2004). This work thus far has not studied how interpersonal judgments could foster cooperation in ways besides punishment and deterrence.

The interpersonal consequences of moral judgments have also been largely neglected in contemporary work in moral psychology. As Haidt and Kesebir (2010) noted in a recent survey, the bulk of the literature has focused on “ethical quandaries.”¹ For instance, research on the often-studied “trolley problem” addresses how people reason in moral dilemmas where deontology and consequentialism prescribe different courses of action. Although this research has yielded a wealth of insights into moral reasoning, we still know comparatively little about the questions that animated early sociological work on morality, such as how interpersonal moral judgments affect the behaviors of those involved.

Below we explore various explanations for the dearth of contemporary work on moral judgments, arguing that it is important to revisit the link between judgments and prosocial behavior. Thereafter, we develop a theoretical account of the social effect of moral judgments. We argue that those who engage in moral judgments subsequently perceive themselves to be more moral and will therefore behave in more trustworthy ways.

We also address the effects of moral judgments on observers. As Goffman (1959) noted, observers tend to assume that actors’ behaviors or performances reflect underlying identities, and, more generally, research finds that people tend to expect behavioral consistency in others (Jones 1990). As a consequence, moral judges should be perceived as more prosocial and trusted to act more prosocially, being, for example, more likely to be chosen as exchange partners under situations of risk and uncertainty.² Thus, our arguments detail two hidden paths through which moral judgments promote cooperation, via the enhanced *trustworthiness* of moral judges and the greater *trust* observers place in moral judges.

We elaborate this argument below and then present the results of three experiments designed to test the predicted effects of moral judgments on the behaviors (Experiment 1) and self-perceptions (Experiment 2) of judges, and the perceptions and behaviors of audiences to judgments (Experiment 3).

Trust, Trustworthiness and Cooperation

When and why people set aside narrow self-interest to behave in cooperative ways, working with others to achieve collective ends, is a longstanding and fundamental puzzle in the social sciences (Kollock 1998). Sustaining cooperation is problematic because it is often in the interest of either or both parties in an interaction to exploit the other for egoistic gain. And because of the risk that one's cooperation may be exploited, it is often prudent for one or both parties to withhold trust. Without trust, cooperation never gets off the ground; without trustworthiness, the potential gains from cooperation cannot be realized. Thus, trust and trustworthiness are critical foundations of voluntary cooperation (Gambetta 1990; Hardin 2002).³

Many early theorists viewed issues such as cooperation and prosocial behavior as central to sociology (Comte 1851; Durkheim 1893; Sorokin 1954). Yet since this foundational work, sociologists have largely drifted away from these problems, typically assuming, rather than explaining, human sociality to move on to other questions of interest (Wrong 1961).

But despite sociologists' reduced attention to the problem of cooperation, researchers outside of sociology have continued to focus on such fundamental questions, most often drawing on theories of self-interest to explain cooperation. For instance, recent research shows that material sanctions, including punishments for noncooperation and rewards for cooperation, produce high levels of cooperation, even among would-be free riders (Yamagishi 1986; Fehr and Gächter 2002). Similarly, research on reputations finds that group members act generously when it leads to a prosocial reputation and, consequently, downstream material and social benefits (Milinski, Semmann and Krambeck 2002).

Although these explanations help illuminate important features of human cooperation, they overlook significant insights from sociological theorizing about how prosocial behavior can emerge via processes besides self-interest. As noted above, foundational sociologists viewed moral judgments as critical to cooperation and prosocial behavior. For instance, Durkheim (1893) and others working in the sociology of deviance (e.g., Erikson 1966) argue that the moral judgments of deviant group members clarify moral boundaries, indicating what is acceptable, normative behavior. These perceptions, in turn, serve to promote adherence to norms and increased prosociality of group members.

Tocqueville reached a similar conclusion in his work on how modern civil arrangements promote prosociality. In *Democracy in America*, he argued that the deliberations of juries affect not just the judged but also the jurors themselves. The jury "serves to communicate the spirit of the judges to the minds of all the citizens. . . . It imbues all classes with a respect for the thing judged, and with the notion of right. . . . By obliging men to turn their attention to affairs which are not exclusively their own, it rubs off that individual egotism which is the rust of society" ([1835] 2002:226).

In contrast to these classical texts, contemporary views cast a more critical eye on moral judgments. Popular culture is replete with examples of moral judges who gain notoriety as a result of their failure to live up to the moral standards by

which they judge others. Moreover, the sanctioning of others is often used as a strategic self-presentation device (Willer, Kuwabara and Macy 2009), which can produce skepticism toward moral exemplifiers. Finally, the lack of research on moral judgments in contemporary sociology may stem partly from the fact that earlier work was associated with theoretical perspectives, such as functionalism, that have largely fallen out of favor in sociology due to their “over-socialized” (Wrong 1961) conceptions of behavior.

Theory and Hypotheses

This section develops an account of how moral judgments impact cooperation, including trust and trustworthiness. Consistent with the classical theorizing just reviewed, we primarily focus on judgments of others’ immoral behaviors in domains around which there is consensus about what is moral versus immoral, e.g., cheating or abusing another’s trust. As explained below, we expect that judgments of immoral behaviors will have larger effects on self-perceptions of morality than judgments of moral behaviors.

In limiting our focus to issues where people agree on what is moral or immoral, we ignore domains (e.g., abortion) characterized by opposing views about what is right or wrong. As we explain in the Discussion section, the effects of judgments in morally contested domains likely depend on whether the observer agrees with the moral judgment or not. Focusing on domains where there is moral consensus allows us to develop an account of the basic mechanisms through which moral judgments affect perceptions and behaviors, irrespective of individual differences in beliefs or moralities. Gaining insight into these processes is important because widespread agreement about what is morally right does not guarantee moral behavior. Indeed, cooperation is problematic precisely because though most people believe that it is wrong to exploit another’s trust for egoistic gain, some nevertheless succumb to the temptation. Moreover, people often suspect that others will give in to this temptation and thus withhold trust.

Moral Judgments and Trustworthiness

Moral judgments do not occur in a social vacuum. Like other social judgments, they are comparative (Kelley 1971). When an actor makes a negative moral evaluation of another, she is taking a stance or making a claim about what she would have done in that type of situation. Moreover, moral judgments imply that the judge accepts the violated norm or moral principle. A central claim of our argument is that the “stance taking” that occurs in the course of morally judging others has important downstream effects on the judge’s own behavior.

That stance taking acts shapes future behavior is based on several central principles from both sociological and psychological social psychology. First, people typically hew closely to a principle of “commitment and consistency.” As Cialdini (2009) notes, “Once we make a choice or take a stand, we will encounter personal and interpersonal pressures to behave consistently with that commitment.” (51) Stance taking might result from taking pledges (Bearman and

Bruckner 2001), making promises (Kerr 1995), prior behaviors (Freedman and Fraser 1966), self-presentations (Goffman 1959) or, in the current framework, making moral judgments of others. We contend that these actions act as commitments that bind a person to consistent behaviors.

The link between former and subsequent behaviors also is also likely driven in part by self-perception processes (Bem 1972; Burger 1999; Gneezy et al. 2012), or a tendency to make attributions about our preferences and identities based on observation of our own behaviors. From this perspective, our behavior may be driven by personal preferences or salient identities, but they might also stem from situational constraints or contextual factors. Whatever the cause, having acted in some way, we infer that we are the sort of person who would be motivated to behave that way.

One of the most powerful demonstrations of the effect of prior commitments on subsequent behaviors is Freedman and Fraser's (1966) classic study of the "foot in the door technique." In their study, homeowners who first agreed to a modest request of displaying a small placard in front of their homes were more likely than a control group to agree to let a different researcher place a much larger, more obstructive sign on their lawns about 2 weeks after the initial, smaller request. Part of the effectiveness of securing the initial commitment is because of changes in self-perception (Cialdini 2009). As Freedman and Fraser explain, upon agreeing to the initial request, the person "may become, in his own eyes, the kind of person who does this sort of thing, who agrees to requests made by strangers, who takes action on things he believes in, who cooperates with good causes." (201)

The prior commitments considered here, and illustrated by the Freedman and Fraser study, differ from the "commitment devices" (Schelling 1960) studied by economists and rational choice theorists. In contrast to rational egoist approaches, our hypothesis posits nonstrategic (unintended) downstream consequences of commitments. As such, we argue that prior behaviors affect subsequent behaviors, even when those behaviors are not observed by others (Pallak, Cook and Sullivan 1980; Kerr 1995; Cialdini 2009).

So far we have focused only on judgments of others' *immoral* behaviors because we expect that, compared with moral praise, moral condemnation will be seen as more diagnostic of a judge's disposition and thus will have larger effects on self-perception and social perception. Past research (Ybarra 2002) finds that negatively valenced acts (aggression, criticism) are believed to speak more to an individual's traits and character than positively valenced acts (warmth, compliments). This may emerge in part because general social expectations to be friendly and agreeable in interaction lead negatively valenced social acts to offer more discerning insight on an individual's personality. Thus, in the context of our research, we expect negative moral judgments of others' immoral acts to engender stronger perceptions of the judge as moral than positive moral praise for moral acts. This argument is consistent with Wiltermuth, Monin and Chow's (2010) demonstration that the tendency to condemn immoral acts is more strongly correlated with the centrality of moral identity than is the tendency to praise moral acts.

Summing up, we expect that moral judgments will lead judges to perceive themselves as more moral and, as a consequence, act more morally in subsequent interactions. Further, we expect that those effects will be limited to judgments of others' immoral behaviors.

- *Hypothesis 1:* After making moral judgments of others' immoral behaviors, individuals will perceive themselves as more moral and act in a more trustworthy way in subsequent anonymous interactions.

We test the behavioral component of Hypothesis 1 in Study 1. Our second experiment addresses the underlying mechanism by assessing the effects of moral judgments on self-perceptions. Study 2 also addresses whether the effects of judgments on self-perceptions are limited to moral condemnation.

Moral Judgments and Trust

In addition to the effect of moral judgments on the judges themselves, we also explore whether others perceive moral judges as trustworthy. The arguments outlined earlier suggest that observers will typically expect that those who make negative moral evaluations of others are not "all talk." Instead, people tend to assume that others will act in line with the identities they claim (Goffman 1959). Generally, when people witness another engaging in a behavior or taking a stand, they assume this reflects a corresponding set of underlying dispositions or values (Jones 1990). Thus, when a person witnesses someone take a moral stance on an issue, by making a moral judgment of another who has acted immorally, the person will tend to expect that the judge is moral and trustworthy. As a consequence, we predict that group members will trust and preferentially associate with those who have made moral judgments in interactional settings with relevant moral content. That is, we know from recent work (e.g., Wang, Suri and Watts 2012) that knowledge of prospective partners' history of cooperative or noncooperative behaviors will affect partner selection and generate higher overall levels of cooperation. We predict that partner choice is also affected by the interpersonal moral judgments potential exchange partners have made.

- *Hypothesis 2a:* Observers will perceive moral judges as more moral and trustworthy.
- *Hypothesis 2b:* Observers will preferentially associate with moral judges under conditions of risk and uncertainty.

As noted earlier, we expect that moral judgments will increase trust only when those judgments are aligned with the norms and values accepted by observers. In moral domains with low consensus, we expect observers are likely only to trust those with similar views.

Study 1: Moral Judgments Increase Judges' Trustworthiness

We conducted three experiments to assess the causal effect of moral judgments on cooperation. Our methods allow us to distinguish our hypothesis from a

weaker claim, which views the link between moral evaluations and subsequent moral behavior as based in strategic reputational maintenance. For instance, a person might use a moral judgment to signal to observers that he is moral. Similarly, upon making a moral judgment, the judge may thereafter act more morally to avoid being labeled a hypocrite. Study 1 rules out such processes by randomly assigning participants to make moral evaluations of others in private (or not), and then measuring their prosociality in a subsequent anonymous context. We predict that those who make private moral evaluations of others will act more trustworthy in a subsequent private interaction.

Methods

Design

Participants were recruited from introductory classrooms at the University of South Carolina for the opportunity to earn money. Seventy-five participants (49% female) took part in the study. Our dependent measure was the standard behavioral measure of trustworthiness (Berg, Dickhaut and McCabe 1995; Buchan, Croson and Dawes 2002). Drawing on methods used in prior work on morality (Batson et al. 1997), we manipulated whether participants witnessed another person act immorally and, if so, whether the participant made a moral evaluation of the immoral act.

Procedure

Participants were scheduled in groups of four to six. Upon arrival, participants were escorted to a private subject room where they were presented with instructions that they would not see other participants at any point during or after the study and that they would be identified only via anonymous participant identifiers.

A number of additional precautions were taken to increase participants' perceptions that their decisions were anonymous and confidential. First, all instructions and procedures were computerized. Thus, there was no contact between the participant and the research assistant from the point at which the research assistant collected the consent form until the debriefing session.

Second, the instructions emphasized that the research assistant would be unaware of the condition to which the participants were assigned.⁴ Specifically, as detailed below, we presented participants with a cover for observing others' immoral actions by noting that a key goal of the research was to refine methods for conducting unbiased studies. The instructions explained that bias can occur when research assistants know to which role each participant has been assigned, and that the researchers were implementing various methods to overcome such bias: the "Computer Assignment Method" had the computer randomly assign participants to roles without the research assistant being aware of which participant was assigned to which condition; the "Participant Self-Assignment Method" ostensibly had a subset of participants assign themselves and others to different conditions or tasks, again without the research assistant knowing the condition assignments.

Participants were told that two studies were being conducted simultaneously: the one for which the participant signed up (“Decision Making”) and an additional study (“Response Consequences”). The instructions explained that participants in the Decision Making study (which included the actual participant) would observe, via computer, a participant in the Response Consequences study, ostensibly to help the researchers evaluate the effectiveness of the Participant Self-Assignment Method. As explained below, this allowed us to expose participants to another’s immoral actions.

Trust Dilemma

Our measure of trustworthiness is based on behavior in the trust dilemma, a cooperation problem that allows researchers to empirically distinguish trust and trustworthiness (Buchan, Croson and Dawes 2002). Prior to the manipulation, participants read instructions for the Decision Making component of the study, including details of the dependent measure. The instructions explained that the (actual) participant and one other participant, whom she would not meet, would interact via computer, and that each would occupy one of two roles (Sender or Receiver) as follows:

The Sender will be given an e-coupon worth \$10. The Sender has two options: Keep the \$10 for himself/herself, or “send” it to the Receiver. If the Sender sends the \$10 to the Receiver, it will be tripled. That is, it will be worth \$30. The Receiver will then decide how much of that \$30 (if any) to return to the Sender. The Receiver can send any amount back: from \$0 to \$30.

Instructions emphasized that the Sender’s and Receiver’s pay for the study depended entirely on whether the Sender sent the e-coupon and, if she did, how much of the resulting \$30 the Receiver returned versus kept for himself.

Participants then completed a comprehension quiz that assessed their understanding of the instructions. Thereafter, the participant was notified that the computer would randomly assign her to the Sender or Receiver position later in the study. In reality, as explained below, all participants were assigned to the Receiver position.

Moral Judgment Task

Shortly after completing the instructions for the dependent measure, the participant received a computerized message stating that others were ready to begin the “Participant Self Assignment Method” process. Next, the participant’s computer appeared to “connect” to the computer of “Participant A.” At that point, A’s computer screen displayed instructions for the “response consequences” task.⁵

Presented as a study of “the effects of positive and neutral response consequences on feelings and reactions,” the instructions explained that participants in the positive consequences task would have a good chance of winning a valuable gift certificate, whereas those in the neutral consequences task

would not (Batson et al. 1997). The explanations of the two tasks concluded by noting:

Most participants find the positive consequence task more exciting than the neutral consequences task. But it is important for our research project to collect information from both the positive and neutral consequences task.

The instructions to ostensible Participant A continued:

In order for the research assistant to remain unaware of which tasks you and Participant D are involved with, you will assign yourself and Participant D to either the Positive (raffle tickets) or Neutral (no raffle tickets) consequences task. Most participants feel that giving both people an equal chance—by, for example, flipping a coin—is the fairest way to assign themselves and the other participant to the tasks. You will be able to use the computer to generate a “computerized coin toss” if you wish. Or you may simply assign either yourself or Participant D to the Positive Consequences Task. The decision is entirely up to you. The other participant does not and will not know that you are assigning tasks; he or she will think that the task assignment was purely by chance. Because of this and because the two of you will never meet, your anonymity is assured.

After appearing to click through each instruction screen, A was then prompted to assign him/herself or Participant D to the positive consequences task, or to let the computer decide.

Experimental Conditions

Participants were randomly assigned to one of three conditions. Each participant either did or did not see the ostensible other assign him/herself to the positive consequences task, an action that prior work (Batson et al. 1997) and pretesting suggested that participants would consider unfair. Those who saw the ostensible other assign him/herself to the positive consequences task either did or did not make a moral judgment of the participant. Thus, because one cannot make a moral judgment of an unknown action, our design does not fully cross witnessing an immoral action (or not) and making an evaluation of it.

Participants in the *no witness* condition were disconnected from the ostensible other’s computer screen prior to the other making a choice. Thereafter, they proceeded to an “evaluation” of the Participant Self-Assignment Method, where they answered innocuous questions about the procedures and instructions. Participants in the *witness immoral action/no judgment* condition answered the same questions as those in the control condition, as well as a manipulation check question: whether A assigned the positive consequences task to self, other or let the computer assign the tasks. Finally, participants in the *moral judgment* condition also answered three Likert-scale moral judgment items about the “fairness,” “selfishness” and “kindness” of A’s decision. The items and response categories were the products of pretests aimed at ensuring that all participants

in the moral judgment condition would, in fact, make moral judgments. As expected, for each item, participants made evaluations significantly beyond the midpoint of the scales (all $ps < .001$).

Dependent Measure

After several moments, the participant's computer transitioned to the Decision Making study. At that point, the computer "randomly" assigned the participant to the trustee ("Receiver") position and an ostensible other participant to the trustor ("Sender") position. The instructions explained that the sender would be given time to decide whether or not to entrust ("invest") the \$10 with the participant. The participant was invariably informed via computer that the other decided to invest the \$10, so that the participant received \$30. The participant was then asked to indicate how much, if any, of these resources to return to the Sender. The amount returned, the standard measure of trustworthiness in the literature, is our dependent measure. Finally, participants were paid, assessed for suspicion and thoroughly debriefed.

Results⁶

Planned pairwise comparisons show that those who judged immoral behaviors were significantly more trustworthy, returning more money in the trust dilemma (mean [M] = 14.96) than those who did not witness the other act immorally (12.38, $t = 2.12$, $p < .05$). Thus, morally judging others led participants to act more morally in a subsequent anonymous interaction. But this comparison leaves open the possibility that simply observing the immoral action, rather than morally judging it, created differences between conditions. A comparison of the treatment with the second control condition reveals that among those who witnessed the other act immorally, those who made moral judgments subsequently acted in a significantly more trustworthy way (14.96) than did those who observed the unfair action, but did not make moral judgments (11.88, $t = 2.91$, $p < .01$).⁷ Participants who made moral judgments returned approximately 21% more of the entrusted endowment than those in the control condition and 26% more than those who saw the other person act unfairly but did not make a moral judgment.

Study 2: Moral Judgments Increase Self-Perceptions of Morality

The results of Study 1 support our argument linking moral judgments to more moral behavior in subsequent interactions. But they are silent on what underlying process drove the effects. We have argued that moral judgments lead individuals to perceive themselves as more moral, but prior work has demonstrated only the converse; that those who see themselves as more moral are more likely to make moral judgments of others (Wiltermuth, Monin and Chow 2010). We thus conducted a follow-up experiment to assess the effects of moral judgments

on a standard measure of moral identity (Aquino and Reed 2002; Stets and Carter 2012). This study also permits a test of our argument that the effects of moral judgments on self-perceptions will be limited to judgments of immoral versus moral behaviors.

Finally, we wanted to rule out an alternative explanation for the Study 1 results. Specifically, it is possible that simply being asked about issues of morality might have primed participants in the moral judgment condition to think about morality and thus to act more morally. Our second study had participants in all conditions recall morally relevant content to rule out this alternative explanation.

Design and Procedure

We recruited participants via Amazon's Mechanical Turk (see Buhrmester, Kwang and Gosling 2011). One hundred and forty-four participants located in the United States (82 females) completed the study online in exchange for payment.

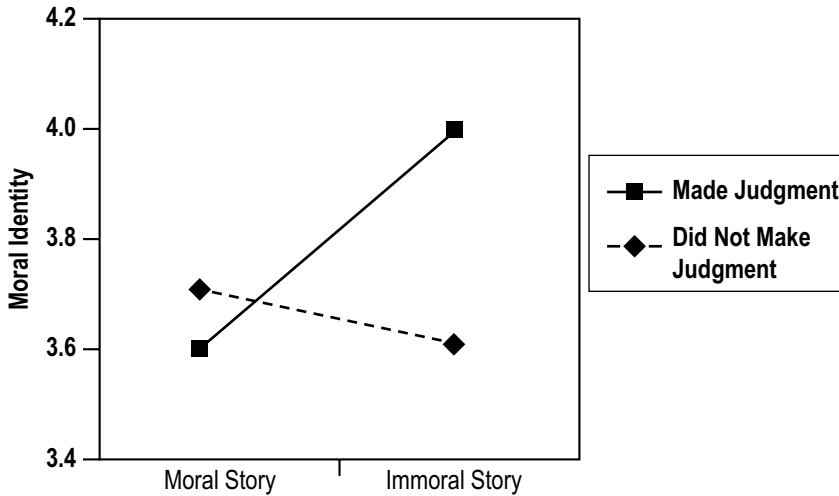
The experiment comprised a 2 (recall of another's moral or immoral act) \times 2 (moral judgment or not) between-subjects design. Participants wrote about a time in the recent past when they witnessed someone behave *morally* or *immorally*, depending on condition, towards a third party.⁸ Participants in the no-judgment condition then continued to the next portion of the study, whereas those in the judgment condition moved on to a series of follow-up questions, where they were asked to further describe the central actor's *morality/immorality*, *selfishness/generosity* and *fairness/unfairness*.

Finally, participants responded to a 10-item moral identity scale (Aquino and Reed 2002), our dependent measure. Because all participants recalled a moral or immoral actor, participants in all conditions should be primed with moral cognitions. But our arguments, outlined earlier, make a more specific prediction: participants who recalled another's immoral behavior and morally judged it will score higher on the moral identity scale than participants in the other three conditions.

Results

Prior to conducting analyses, 22 respondents were removed from the sample for not following instructions for the story task (i.e., they either wrote about their own immoral/moral behaviors or about behaviors that benefited/harmed themselves). Thus, our analyses are based on 122 participants (72 females).

Responses to the moral identity scale were highly reliable ($\alpha = .80$); therefore, items were averaged into a single scale, where higher values indicate greater importance of one's moral identity. Next, we conducted a 2 (immoral vs. moral) \times 2 (judgment vs. no judgment) analysis of variance (ANOVA). Results revealed the predicted interaction of immoral stories and judgment, $F(1, 118) = 5.08$, $p = .03$. As shown in Figure 1, participants who recalled an immoral act *and* condemned it scored higher on the moral identity scale ($M = 3.97$, standard deviation [SD] = .46) than those who wrote about an immoral act but did not judge it ($M = 3.63$, $SD = .58$) or those who wrote about a moral act and judged

Figure 1. Mean Score on Moral Identity Scale by Condition

it ($M = 3.61$, $SD = .60$) or did not judge it ($M = 3.71$, $SD = .50$). There were no main effects for story valence, $F(1, 118) = 2.13$, $p = .15$, or for making judgments, $F(1, 118) = 1.60$, $p = .21$.

The results of this follow-up study are important for a number of reasons. First, they suggest that the effects of judgments of immoral acts are not due to priming effects and thus help rule out this alternative explanation of the Study 1 findings. Second, they extend our findings by linking judgments to self-perceptions of moral identity, a strong predictor of a wide range of moral and prosocial behaviors (Aquino and Reed 2002; Stets and Carter 2012). Finally, by showing that self-perceptions of moral identity are affected by moral condemnation but not moral praise, they provide a clearer insight on what types of moral judgments affect moral identity and prosociality.

Study 3: Observers Trust Moral Judges

We now turn to a test of our second hypothesis. As discussed earlier, people not only tend to exhibit behavioral consistency themselves, but they also expect it in others. Study 3 was designed to address the effects of moral judgments on the perceptions (Hypothesis 2a) and behaviors (Hypothesis 2b) of fellow group members. We predict that observers to moral judgments will view judges as more moral and trustworthy, and will tend to preferentially associate with them in social exchanges involving risk and uncertainty.

Methods

Design

Participants were recruited from introductory classes for the opportunity to earn money. A total of 69 participants (62% female) took part in Study 3.

Participants were exposed to two targets, one of whom made stronger moral judgments about a person involved in a plagiarism case. Our primary dependent measure was whether participants were more apt to select the person who makes the stronger moral judgment as an exchange partner, a key indicator of trust (Kollock 1994). In addition, participants completed a second behavioral measure of trust, as well as items measuring their perceptions of the targets' morality and trustworthiness.

Procedures

Participants were scheduled in groups of four to five. Upon arrival to the laboratory, each participant was escorted to a private subject room. Participants were assured that they would not meet other participants at any point during or after the study, and that participants would be identified only via anonymous identifiers.

The experiment was framed to participants as involving two studies, the first of which addressed university students' perceptions of academic dishonesty. Instructions explained that there were four participants taking part in the study, and each participant would be randomly assigned to one of two roles. Two "Raters" would read a description of a case heard by the university's Office of Academic Integrity and answer survey questions about the case. After reading the case description themselves, two "Comparers" would view the raters' responses and estimate the extent to which the raters' perceptions of academic dishonesty were in agreement. The ostensible purpose of this exercise was to give researchers an estimate of the typical student's attitudes about academic dishonesty. This task provided a cover for exposing participants to both a "stronger" and a "weaker" moral judge. In reality, all participants served as Comparers and compared the judgments of two ostensible others.

The ratings participants compared contained responses to several questions about perceived frequency of academic dishonesty, as well as the raters' responses to questions about the dishonest person's ("Mark") morality. One ostensible Rater, the strong moral judge, rated Mark on a 6-point Likert scale (i.e., from "very moral" to "very immoral") as "very" immoral, selfish, unfair and dishonest. The weak moral judge rated Mark as "slightly" immoral, selfish, unfair and dishonest.

Embedded among questions relevant to the cover story—e.g., the level of agreement between raters on several issues—participants reported their perceptions of the raters' morality and trustworthiness on 6-point Likert scales. These are our first two dependent measures.

Behavioral Measures

Participants then made a decision in the trust dilemma like the one from Study 1. Study 3 participants occupied the role of trustors. Instructions explained that, for administrative convenience, the two raters would occupy the trustee ("Receiver") roles and the comparers (thus, the actual participant) would fill the two trustor ("Sender") roles. The participant was asked to select one of the two raters as an

interaction partner for the trust dilemma. Whether the participant chose the strong or weak judge was our primary dependent measure of trust (Kollock 1994).

Finally, the instructions informed participants that in the event that both trustors chose the same trustee, partners would be randomly matched. It was therefore important for the participant to indicate investment amounts for each potential partner. Thus, in addition to our primary dependent measure, we measured whether participants would entrust a greater portion of their \$10 endowment to the strong versus weak judge. As in Study 1, any amount sent by the trustor was tripled. The trustee would then decide how much of the tripled amount, if any, to return to the sender. Because participants ostensibly risked losing their entire endowment to an untrustworthy partner, higher investments provide an additional behavioral measure of perceived morality and trustworthiness. Trustor's behavior in the trust dilemma is arguably the most widely used behavioral measure of trust, and complements our measure of trustworthiness from Study 1.

Results

Results for our primary dependent measure, whether participants were more likely to pick the strong moral judge as an exchange partner, strongly support Hypothesis 2b; the majority of participants (75%) picked the strong judge as an exchange partner, a rate which is significantly greater than chance, $\chi^2 = 17.76$, $p < .001$. Results of our second behavioral measure also show that participants entrusted significantly higher proportions of their endowments to the strong judge ($M = 5.90$) than to the weaker judge ($M = 4.36$, $t = 4.26$, $p < .001$).

Finally, as predicted (Hypothesis 2a), participants rated the strong judge as significantly more moral (6.74 vs. 4.97, $t = 14.45$, $p < .001$) and trustworthy (6.48 vs. 5.06, $t = 9.16$, $p < .001$) than the weak judge. In short, the four dependent measures offer strong and consistent support for our prediction that in those moral domains characterized by consensus, individuals who make strong moral judgments are trusted more and viewed as preferable exchange partners in interactional settings with moral content, like the trust dilemma.

Follow-Up Study

Although the results of Study 3 are consistent with our hypothesis that strong moral judges are viewed as more moral and trustworthy, alternative explanations for the findings are possible. First, rather than trusting stronger moral judges, people may instead trust those who make judgments similar to those that they would have made themselves. If participants' own evaluations of the plagiarism case were more like those of the strong than the weak judge, then this greater similarity might have led to trust in the strong judge. Second, if participants viewed the plagiarist as highly immoral, the failure of the weaker judge to strongly condemn him might have led participants to view the weak judge as weird or amoral. Thus, the Study 3 findings might have been driven more by distrust of the weak judge than trust of the strong judge.

We conducted a vignette-based experiment ($N = 79$) to distinguish our hypothesis from these alternatives. Respondents, sampled from the same population as

Study 3, first evaluated the same case of academic misconduct as the judges in Study 3. They rated the scenario target's morality, selfishness, honesty and fairness on 6-point Likert scales.⁹ After giving their ratings, participants viewed the moral judgments of three ostensible others, one with evaluations identical to the strong judge in Study 3, one with evaluations identical to the weak judge's, and a third set of evaluations from an intermediate judge, whose evaluations fell exactly in between. As in Study 3, participants rated each judge's morality and trustworthiness.

Participants' ratings were far more similar to the intermediate judge than either the strong ($t = 7.07, p < .001$) or the weak judge ($t = -4.84, p < .001$). Their ratings fell between those of the intermediate and weak judges' and thus were (non-significantly) more similar to the evaluations of the *weak* judge than the strong judge, $p = .48$. But despite this similarity to the intermediate judge, participants consistently rated the strong judge as both more moral and trustworthy than either the intermediate or weak judge (for all comparisons, $t \geq 3.37; p \leq .001$).

To more explicitly address the relative impact of similarity and judgment strength on participants' perceptions of judges' morality and trustworthiness, we submitted each dependent measure to a repeated-measures ANOVA. Each model included moral judgment strength as a within-subjects factor and participants' own ratings of the plagiarism case as a covariate. Both models revealed strong main effects of moral judgment strength ($F > 8.75; p < .001$). But participants' own ratings of the case had only a marginal effect on perceptions of the judges' morality ($F = 3.92, p = .051$) and no effect on perceptions of the judges' trustworthiness ($F = 2.61, p = .11$). The results therefore provide clear evidence that the greater perceived morality and trustworthiness of strong moral judges in Study 3 was driven by the strength of those judgments, rather than similarity with the judgments participants would have made themselves. Further, these results suggest it is unlikely that participants in Study 3 simply viewed the weak judge as weird or amoral since the participants in the vignette study reported they would have made judgments that were, if anything, somewhat closer to those of the weak judge than the strong judge.

Discussion and Conclusion

How societies reduce conflicts between individual and collective interests and promote prosocial behavior was a question central to classical sociological thinking (Comte 1851; Durkheim 1893). Yet since this foundational work sociology has largely moved on to other questions, assuming rather than explaining human prosociality (Wrong 1961). As a consequence, most explanations come from outside sociology. While these explanations have yielded many important insights, we have argued that our understanding of prosociality could be greatly enhanced by revisiting a classical sociological insight on the problem of social order: the *moral judgments* that group members make about one another serve to attenuate conflicts between individual and collective interests.

Our research has identified two paths through which moral judgments increase prosociality: via the enhanced *trustworthiness* of moral judges and the

greater *trust* observers place in moral judges. We discuss each of these paths in turn.

Our first two studies demonstrated the downstream effects of moral judgments on moral identity (Study 2) and heightened prosociality (Study 1). On the surface, the results of these studies might appear to run counter to the *moral licensing* literature, which suggests that moral behavior in one context “licenses” one to act less morally in subsequent settings (e.g., Sachdeva, Iliev and Medin 2009). If this is the case, wouldn’t moral judgments *decrease* judges’ subsequent prosociality? Recent moral psychology work has sought to reconcile similar tensions between past research on moral licensing and moral identity, arguing that the level of abstraction at which one construes one’s own behavior—either concrete or abstract—serves as a critical moderating variable (Conway and Peetz 2012). Moral acts that provide direct benefits to others, such as assisting another or giving to a charity, are construed at a concrete level immediately after the act and lead to reduced moral behavior subsequently because they produce a sense that one has “done enough” for others at that time.

Other morally-relevant behavior—such as recalling moral acts from the distant past—are relevant to one’s moral identity via more abstract processing and tend to activate or enhance one’s moral identity, leading to greater subsequent prosocial behavior. Here, we argue that individuals who have made moral judgments of others perceive themselves as taking a stance in support of an abstract moral principle, but do not view themselves as having provided direct benefits to others. Consequently, moral judgments are likely to be construed as relevant to one’s own moral identity at an abstract rather than a concrete level (Eyal, Liberman and Trope 2008). The effect of this abstract construal is to *enhance* feelings of moral identity and foster greater prosocial behaviors, rather than giving rise to moral licensing and a feeling that one has “done enough.” Our results are consistent with this interpretation, though future work should more fully integrate the study of moral judgments with this emerging, general understanding of moral identity and licensing effects.

Findings from our first two studies complement recent sociological work that takes moral identity as a starting point to explain variation in moral behavior (Stets and Carter 2012). Linking these two lines of research suggest a path through which moral behaviors and moral identity might mutually reinforce one another: the interpersonal moral judgments actors make about others lead to heightened self-perceptions of moral identity (Study 2). Greater self-perceived morality, in turn, has been shown to generate an array of moral behaviors (Aquino and Reed 2002; Stets and Carter 2012; see also Study 1, above). Here, we found that even a single expression of a moral judgment of another was sufficient to shift an individual’s moral self-perception and related subsequent behaviors, findings that underscore the power of moral identity processes.

Not only do moral judgments increase the self-perceived morality and prosocial behavior of judges, our third study further demonstrates that observers anticipate these effects. Strong moral judges were perceived as more moral and trustworthy. In addition, they were preferentially chosen as interaction partners and trusted to a greater extent than weak judges. These findings suggest that

moral judgments may be socially beneficial because they establish a basis of interpersonal trust. Trust has been linked to a range of individual- and societal-level benefits (Coleman 1990; Fukuyama 1995; Putnam 2001) and is considered fundamental to social life. But the benefits of trust can be realized only if trust is honored rather than exploited (Gambetta 1990). It is therefore important that Study 1 showed that moral judgments also establish a basis of trustworthiness. Considered together, the results from these experiments suggest that by promoting trust and trustworthiness, moral judgments can help resolve conflicts between individual and collective interests, thus increasing overall welfare and promoting social order. More generally, by highlighting two interrelated paths through which moral judgments increase prosociality, our arguments and findings underscore the emphasis that classical sociological theorists placed on moral judgments and highlight the need for greater attention to moral judgments in future work on prosociality.

Future Directions

The arguments and evidence presented above suggest an array of new directions for future work on moral judgments. Most immediately, future work should explore likely moderators of moral judgments. We limited our focus to those domains where there is considerable consensus about what is moral or immoral. Under these conditions, Study 3 found that strong judges are trusted more and viewed as more moral than weaker judges. But some domains (e.g., abortion or same-sex marriage) are characterized by fundamentally opposing views of right and wrong. Indeed, such domains often coincide with rival “moral communities” (Durkheim 1912) with distinct moral norms and well-defined intergroup boundaries. Researchers have long noted that intergroup boundaries have powerful effects on interpersonal trust (Brewer 1999). Thus, in domains where moral consensus is low or where moral norms vary along ingroup/outgroup lines, the perceived trustworthiness of a moral judge will likely be moderated by whether the observer agrees with the moral judgment, or whether the judge is an ingroup or outgroup member.

As suggested earlier, our finding that moral judges are preferentially interacted with offers one path to solving cooperation problems (e.g., Wang, Suri and Watts 2012), as moral judgments lead to disproportionate interaction with more cooperative people. Additionally, if moral judgments lead individuals to have more extensive network connections and more positive reputations, then it seems likely that those who make judgments will often ascend to positions of leadership, status and influence in groups. This dynamic would have important implications for individual and group outcomes. First it is possible that such reputational benefits could trigger a dynamic of “competitive moral judgmentalism” (see Barclay and Willer 2007), as individuals vie to be more judgmental than one another in pursuit of reputational benefits, a dynamic that might be limited only by whatever aversion to extreme judgmentalism group members harbor. This dynamic would be most likely in communities based on strong moral principles. Further, the ascension of more judgmental individuals to positions of

moral authority and influence as a result of their apparent trustworthiness and integrity could, ironically, establish the preconditions for conspicuous falls from grace given the potential for power to corrupt (Keltner, Gruenfeld and Anderson 2003), which could overwhelm the positive effect of moral judgments on moral behavior identified here. These implications could prove fruitful avenues for future research.

Notes

1. Indeed, these literatures use the phrase “moral judgment” in a different way than the more social, or interpersonal, sense in which we use it. In psychology and philosophy, moral judgments typically refer to outcomes of conscious reasoning about moral dilemmas (Piaget 1965), or to “intuitions” about the rightness or wrongness of various courses of action in moral dilemmas (Haidt 2007). Here, we use them to refer to the evaluations (good/bad, fair/unfair, just/unjust) that one person makes about one or more others’ behaviors.
2. As explained more fully below, we focus primarily on moral judgments that occur in those domains where there is consensus about what is moral versus immoral.
3. Our use of these concepts is consistent with prior work. We define trust as “a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another” (Rousseau et al. 1998:395). Trustworthiness, in turn, is defined as positive intentions and behavior, i.e., one acts trustworthily when one does not abuse trust when there is an incentive and opportunity to do so (see Hardin 2002). As these definitions imply, trust and trustworthiness are at issue under situations of social uncertainty (Kollock 1994).
4. These precautions seemed to have increased participants’ feelings of anonymity. Prior work shows that increased anonymity leads to less prosocial behavior (e.g., Barclay and Willer 2007). The level of trustworthiness in the control condition, reported below, was lower than the level observed in a study conducted using comparable procedures in the same laboratory (Simpson and Eriksson 2009), where these additional precautions to increase feelings of anonymity were not taken.
5. The “response consequences” cover story was developed by Batson et al. (1997) as a dependent measure of moral decision making. We retooled the procedure for our manipulation.
6. Five (of seventy-five) participants expressed suspicion about whether the other participants were real. One was in the control condition, and two were in each of the remaining two conditions. In addition, two participants (one in each of the “witness” conditions) failed the manipulation check; one thought the other assigned the task by letting the computer decide and one thought the other assigned the positive consequences task by assigning the positive task to another. Eliminating these seven left a total of 68 participants.
7. We did not have an a priori expectation about differences in the two nonjudgment conditions. While those in the witness/nonjudgment condition were somewhat less trustworthy than those in the no-witness condition, suggesting a potential modeling effect, this difference did not approach statistical significance ($t = .324, p = .75$).
8. Participants were instructed to describe behaviors that benefited or harmed a third party (rather than behaviors that benefited or harmed them personally) to avoid activating feelings of gratitude or retaliation that may be associated with being harmed/benefited (Conway and Peetz 2012).

9. Dropping ratings of *selfishness* resulted in a more reliable scale. Thus, the results reported below are based on a composite of participants' ratings of morality, fairness, and honesty. Analyses using the four-item measure and each single-item measure yielded substantively identical results.

References

- Aquino, Karl, and Americus Reed II. 2002. "The Self-Importance of Moral Identity." *Journal of Personality and Social Psychology* 83:1423-40.
- Barclay, Pat, and Robb Willer. 2007. "Partner Choice Creates Competitive Altruism in Humans." *Proceedings of the Royal Society of London* 274:749-53.
- Batson, C. Daniel, Diane Kobrynowicz, Jessica Dinnerstein, Hannah Kampf, and Angela Wilson. 1997. "In a Very Different Voice: Unmasking Moral Hypocrisy." *Journal of Personality and Social Psychology* 72:1335-48.
- Bearman, Peter S., and Hannah Bruckner. 2001. "Promising the Future: Virginity Pledges and First Intercourse." *American Journal of Sociology* 106:859-912.
- Bem, Daryl J. 1972. "Self Perception Theory." *Advances in Experimental Social Psychology* 6:1-62.
- Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10:122-42.
- Brewer, Marilynn. 1999. "The Psychology of Prejudice: Ingroup Love or Outgroup Hate?" *Journal of Social Issues* 55:429-44.
- Buchan, Nancy, Rachel Croson, and Robyn Dawes. 2002. "Swift Neighbors and Persistent Strangers." *American Journal of Sociology* 108:168-206.
- Buhrmester, Michael, Tracy Kwang, and Samuel Gosling. 2011. "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?" *Perspectives on Psychological Science* 6:3-5.
- Burger, Jerry. 1999. "The Foot-in-the-Door Compliance Procedure: A Multiple-Process Analysis and Review." *Personality and Social Psychology Review* 3:303-25.
- Cialdini, Robert. 2009. *Influence: Science and Practice*. New York, NY: Quill.
- Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge, MA: Harvard University Press.
- Comte, Auguste. [1851] 1973. *System of Positive Polity or Treatise on Sociology Including the Religion of Humanity*. New York, NY: Lenox Hill.
- Conway, Paul, and Johanna Peetz. 2012. "When Does Feeling Moral Actually Make You a Better Person?" *Personality and Social Psychology Bulletin* 38:901-19.
- Durkheim, Emile. [1893] 1984. *The Division of Labor in Society*. New York, NY: Free Press.
- _____. [1912] 1995. *The Elementary Forms of Religious Life*. New York, NY: Free Press.
- Erikson, Kai T. 1966. *Wayward Puritans*. Boston, MA: Allyn & Bacon.
- Eyal, Tal, Nira Liberman, and Yaacov Trope. 2008. "Judging Near and Distant Virtue and Vice." *Journal of Experimental Social Psychology* 44:1204-09.
- Fehr, Ernst, and Simon Gächter. 2002. "Altruistic Punishment in Humans." *Nature* 415:137-40.
- Freedman, Jonathan, and Scott Fraser. 1966. "Compliance Without Pressure: The Foot-in-the-Door Technique." *Journal of Personality and Social Psychology* 37:580-90.
- Fukuyama, Francis. 1995. *Trust: The Social Virtues and Creation of Prosperity*. New York, NY: Free Press.
- Gambetta, Diego. 1990. *Trust: Making and Breaking Cooperative Relations*. Oxford, UK: Blackwell.
- Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. New York, NY: Doubleday.
- Gneezy, Ayelet, Alex Imas, Amber Brown, Leif Nelson, and Michael Norton. 2012. "Paying to be Nice: Consistency and Costly Prosocial Behavior." *Management Science* 58:179-87.
- Haidt, Jonathan. 2007. "The New Synthesis in Moral Psychology." *Science* 316:998-1002.

- Haidt, Jonathan, and Selin Kesebir. 2010. "Morality." Pp. 797-832. *Handbook of Social Psychology*. S. Fiske, D. Gilbert, and G Lindzey, editors. Hoboken, NJ: Wiley.
- Hardin, Russell. 1982. *Collective Action*. Baltimore, MD: Johns Hopkins University Press.
- _____. 2002. *Trust and Trustworthiness*. New York, NY: Russell Sage.
- Hechter, Michael. 1987. *Principles of Group Solidarity*. Berkeley, CA: UC Press.
- Jones, Edward E. 1990. *Interpersonal Perception*. New York, NY: W.H. Freeman Company.
- Kelley, Harold H. 1971. "Moral Evaluation." *American Psychologist* 26:293-300.
- Keltner, Dacher, Deborah H. Gruenfeld, and Cameron Anderson. 2003. "Power, Approach, and Inhibition." *Psychological Review* 110:265-84.
- Kerr, Norbert. 1995. "Norms in Social Dilemmas." In D. Schroeder *Social Dilemmas*. New York, NY: Praeger.
- Kollock, Peter. 1994. "The Emergence of Exchange Structures: An Experimental Study of Uncertainty, Commitment, and Trust." *American Journal of Sociology* 100:313-45.
- _____. 1998. "Social Dilemmas." *Annual Review of Sociology* 24:183-214.
- Milinski, Manfred, Dirk Semmann, and Hans-Jurgen Krambeck. 2002. "Reputation Helps Solve the 'Tragedy of the Commons.'" *Nature* 415:424-26.
- Pallak, Michael, David Cook, and John Sullivan. 1980. "Commitment and Energy Conservation." *Applied Social Psychology Annual* 1:235-53.
- Piaget, Jean. [1932] 1965. *The Moral Judgment of the Child*. New York: Free Press.
- Putnam, Robert. 2001. *Bowling Alone*. New York, NY: Simon & Schuster.
- Rousseau, Denise M., Sim B. Sitkin, Ronald S. Burt, and Colin Camerer. 1998. "Not So Different After All: A Cross-discipline View of Trust." *Academy of Management Review* 23:393-404.
- Sachdeva, Sonya, Rumen Iliev, and Douglas Medin. 2009. "Sinning Saints and Saintly Sinners: The Paradox of Moral Self-Regulation." *Psychological Science* 20:523-28.
- Schelling, Thomas C. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
- Semmann, Dirk, Hans-Jurgen Krambeck, and Manfred Milinski. 2004. "Strategic Investment in Reputation." *Behavioral Ecology and Sociobiology* 56:248-52.
- Simpson, Brent, and Kimmo Eriksson. 2009. "The Dynamics of Contracts and Generalized Trustworthiness." *Rationality and Society* 21:59-80.
- Sorokin, Pitirim. 1954. *The Ways and Powers of Love*. Boston, MA: Beacon Press.
- Stets, Jan, and Michael Carter. 2012. "A Theory of the Self for the Sociology of Morality." *American Sociological Review* 74:192-215.
- de Tocqueville, Alexis. [1835] 2002. *Democracy in America*. Chicago, IL: University of Chicago Press.
- Wang, Jing, Siddharth Suri, and Duncan J. Watts. 2012. "Cooperation and Assortativity with Dynamic Partner Updating." *Proceedings of the National Academy of Sciences* 109: 14363-368.
- Willer, Robb, Ko Kuwabara, and Michael Macy. 2009. "The False Enforcement of Unpopular Norms." *American Journal of Sociology* 115:451-90.
- Wiltermuth, Scott, Benoit Monin, and Rosalind Chow. 2010. "The Orthogonality of Praise and Condemnation in Moral Judgment." *Social Psychological and Personality Science* 1:302-10.
- Wrong, Dennis. 1961. "The Oversocialized Conception of Man in Modern Society." *American Sociological Review* 26:183-93.
- _____. 1994. *The Problem of Order*. New York, NY: Free Press.
- Yamagishi, Toshio. 1986. "The Provision of a Sanctioning System as a Public Good." *Journal of Personality and Social Psychology* 51:110-16.
- Ybarra, Oscar. 2002. "Naïve Causal Understanding of Valenced Behaviors and Its Implications for Social Information Processing." *Psychological Bulletin* 128:421-41.