Note: This PDF contains the following papers:


The Topography of Generosity: Asymmetric Evaluations of Prosocial Actions

Nadav Klein and Nicholas Epley
University of Chicago

Prosociality is considered a virtue. Those who care for others are admired, whereas those who care only for themselves are despised. For one’s reputation, it pays to be nice. Does it pay to be even nicer? Four experiments assess reputational inferences across the entire range of prosocial outcomes in zero-sum interactions, from completely selfish to completely selfless actions. We observed consistent nonlinear evaluations: Participants evaluated selfish actions more negatively than equitable actions, but they did not evaluate selfless actions more favorably than equitable actions. This asymptotic pattern reflected monotonic evaluations for increasingly selfish actions and insensitivity to increasingly selfless actions. It pays to be nice but not to be really nice. Additional experiments suggest that this pattern stems partly from failing to make spontaneous comparisons between varying degrees of selflessness. We suggest that these reputational incentives could guide social norms, encouraging equitable actions but discouraging extremely selfless actions.

Keywords: prosocial behavior, reputation, fairness, social judgment, attribution

Supplemental materials: http://dx.doi.org/10.1037/xge0000025.supp

It is difficult to think of generosity and selfishness without also thinking about praise and scorn, respectively. In philosophy, Descartes, Hume, and Aquinas all identified self-sacrifice for the benefit of others as a major ethical accomplishment in a person’s life. In religion, the Bible exalts generosity as the most important spiritual virtue (Corinthians 1:13) and links selfishness with evil (James 3:16; Romans 2:8). Buddha went further, teaching that compassion (karuna) not merely is a virtue but is one of only two essential qualities for enlightenment (wisdom being the other). In literature, Ebenezer Scrooge becomes beloved only after exchanging his miserly ways for generosity toward others (Dickens, 1843). In modern life, no representative sample is needed to confirm that seemingly self-sacrificing figures like Mother Teresa and Mahatma Gandhi are admired whereas selfish actors like Jeff Skilling and Bernard Madoff are despised.

These examples reflect a widespread tendency to admire prosociality in others and to despise selfishness. A child who shares a bite of his cookie will be liked more than one who shares none of it. A wealthy person who gives 10% of her income to charity will appear more caring than one who gives none of her income. A professor who offers half a workday to help students will be more beloved than a professor who offers none of the day. In terms of one’s reputation, it pays to be nice.

Does it pay to be even nicer? That is, does behaving even more prosocially—benefiting others at an increasing cost to oneself—lead to an even more positive reputation? Is a child who gives away an entire cookie liked more than one who shares only a bite? Is the wealthy person who gives 50% of her income seen as more caring than one who gives away only 5%? Is a professor who gives an entire workday to help students admired more than one who gives only half a day?

These questions address a fundamental issue in social judgment: Does behaving more prosocially lead to a consistently more positive reputation across the entire range of prosocial actions from completely selfish to completely selfless? Existing research typically measures evaluations of only discrete points along this continuum, such as clearly selfish versus clearly generous actions (e.g., Almenberg, Dreber, Apicella, & Rand, 2011; Berman & Small, 2012; Gray, Ward, & Norton, 2014), without studying the entire range of possible behavior. These questions also address a fundamental issue in self-regulation: If a person wants to maintain a positive reputation, how exactly should he or she behave? Some acts of kindness could incur costs that are not repaid in reputational value. Knowing the reputational value of prosocial behavior may explain how people optimize the potential trade-off between reputational benefits and personal costs.

We conducted 11 experiments—eight described here and three in the online supplemental materials—that address these questions. In each experiment, we study evaluations in zero-sum contexts where people can benefit either themselves or others in varying degrees. In these contexts, the precise degree of selfish versus selfless action is maximally clear.
Theoretical Possibilities

If prosociality is considered a virtue, then a person who shows more of it should earn a more positive reputation than a person who shows less. This predicts a relatively monotonic pattern across the range of prosocial actions, such that selfish actors are evaluated more negatively than moderately selfless (or equitable) actors, and extremely selfless actors (who maximally benefit others at a cost to the self) are evaluated more favorably than only moderately selfless actors (see Figure 1). Evaluations of a person’s prosocial actions might therefore resemble judgments of other personal characteristics, such as height. As a person grows taller, the person is judged to be taller across the entire range of human height. At no point across the spectrum does a taller person stop looking taller or a shorter person stop looking shorter. On this account, it pays to be nice and pays more to be really nice.

As logical as this monotonic pattern may be, we believe there are at least two reasons to doubt that people’s reputations actually conform to it. First, others’ selfless behavior may become a threat to one’s own sense of self-esteem, a threat that may combat by derogating others’ selfless actions (Fetchenhauer & Dunning, 2010; Minson & Monin, 2012; Monin, Sawyer, & Marquez, 2008). This predicts an inverted-U pattern, such that selfish actors are evaluated more negatively than equitable actors, but selfish actors who benefit others even more than themselves are evaluated more negatively than merely equitable actors (see Figure 1). On this account, it actually pays less to be really nice.

Second, to value outcomes of varying magnitudes rationally, these outcomes have to be evaluated in comparison to each other (Hsee & Zhang, 2010). People may indeed evaluate extremely prosocial actions more positively than moderately prosocial actions, but if they do not compare one against the other they may evaluate the two outcomes similarly. If evaluating a person who donated 5% of income to charity does not lead people to think of the possibility of giving away 50% to charity, then a mildly charitable person may be evaluated the same as an extremely charitable person when evaluated in isolation. Existing research demonstrates that value sensitivity depends partly on stimulus familiarity (Hsee, Loevenstein, 2010). The more familiar or knowledgeable a person is with a stimulus, the more sensitive evaluations will be to the objective value of the stimulus (Ariely & Loewenstein, 2000; Laming, 1997; Linville, 1982; Parducci, 1965; Stevens, 1975). For example, in one study the duration of a commute affected evaluations of that commute only for people who experienced this commute on a frequent basis (Morewedge, Kassam, Hsee, & Caruso, 2009). In another study, providing participants with distributional information about admissions test scores markedly increased discernment between low- and high-scoring candidates (Hsee, Loewenstein, Blount, & Bazerman, 1999). Likewise, norm theory suggests that when people evaluate familiar stimuli (versus unfamiliar stimuli), they are more likely to spontaneously generate comparison standards (Kahneman & Miller, 1986).

Variations of selfless behavior—in which people benefit others at varying costs to the self—are simply less common in everyday life than selfish behavior (Dawkins, 1976; Trivers, 1971). Americans, for instance, have donated roughly 2% of gross domestic product to charity consistently for the past 30 years while keeping the bulk of the remaining 98% for themselves (Giving USA, 2013). A recent review of dictator games (N = 20,183) found that more than 70% gave less than an even split, whereas only 13.1% gave more than an even split (Engel, 2011). Because selfless actions are relatively rare and unfamiliar, evaluations of selfless actions may be less sensitive to varying levels of generosity, leading them to be judged categorically rather than relatively. In contrast, selfish actions may be more likely to trigger spontaneous comparisons because they are both negative and common, implying that they would be evaluated in a broader perspective, in relation to the magnitude of selfishness. This predicts an asymptotic pattern: People judge increasing degrees of selfishness increasingly more negatively but do not judge increasing degrees of selflessness more positively (see Figure 1). On this account, it pays to be nice but pays no more to be really nice.

Overview of Experiments

We conducted a series of experiments to identify whether evaluations of prosocial actions follow a relatively monotonic, inverted-U, or asymptotic pattern. In each, participants evaluated a person who behaved relatively selfishly (benefiting only the self), selflessly (generously benefiting others at an increasing cost to oneself), or some point in between. Experiment 2 also tests whether people know how their prosocial actions are evaluated in the eyes of others, a crucial judgment for self-regulation. Experiments 4–6 test a possible explanation for the pattern of evaluations we observe.

A person’s reputation may vary along many different dimensions, but existing research demonstrates that it typically varies along only two fundamental dimensions: warmth and competence (Fiske, Cuddy, & Glick, 2007; Judd, James-Hawkins, & Yzerbyt, 2005). Warmth is more dominant in judgment (Willis and Todorov, 2006; Wojciszke & Abele, 2008) and is related to other-oriented outcomes (e.g., friendliness, trustworthiness, morality), whereas competence is secondary and related to self-oriented outcomes (e.g., intelligence, talent, skill). We expected that prosocial actions would affect only reputations of warmth, but we measured competence in some of the studies in case the two were inversely related such that extreme acts of kindness are also seen as a sign of incompetence (Judd et al., 2005).

Figure 1. Theoretical relationships between prosociality and reputational inference.
Experiments 1a and 1b: Increasingly Generous Donations

We first tested our hypotheses in a situation in which people commonly witness and engage in a wide range of prosocial behavior. Specifically, our university orchestra concerts are free but include a suggested donation of $10. Initial conversations with orchestra personnel verified that concertgoers vary in their behavior, with some donating nothing and some donating well above the suggested donation (in our sample, the highest donation was $90 for a concert ticket). We asked actual concertgoers (Experiment 1a) or observers (Experiment 1b) to evaluate the character of a person who behaved increasingly generously (i.e., donated less, exactly, or more than the suggested donation in varying degrees).

Method

Experiment 1a. People who attended two university orchestra concerts received a survey in their program that they could complete voluntarily and return at the exit (N = 102, 57% female, M_age = 47.7 years) The orchestra suggests a $10 donation at the concert, with donation baskets placed at the entrance. Thus, donating at concerts is optional.

The survey asked concertgoers to consider “Tom W.,” a typical classical music lover who attended a recent concert. The survey contained a generic description of Tom and then indicated that he donated $0, $5, $10, $20, or $50 at the concert. Experiment 1a described a behavior in everyday life we cited in the introduction, relatively prosocial actions were more rare than relatively selfish actions.

Impressions. To test our main hypotheses, we first created two composites, one of the warmth items (α = .76) and one of the competence items (α = .81). Nine participants failed to answer at least one of the impression items (5 related to warmth, 6 related to competence). We did not include these participants with missing data when creating our composites. However, including them by simply averaging across the other items does not affect any of the following results in any meaningful way.

Results

Past concert attendance. Participants attended 2.51 concerts on average during the academic year, and 57.2% of participants attended more than one concert.

Donation behavior. Ninety of our participants reported their own donations. A substantial portion (41.1%) donated the suggested $10. Among the rest, 48.9% donated less than $10 (26.7% donated nothing), and 10.0% of participants donated more than $10 (range = $15–$90). Thus, selfish behavior (donating less than the suggested amount) was approximately five times more common than generous behavior (donating more than the suggested amount) in this context. Consistent with the more general pattern of behavior in everyday life we cited in the introduction, relatively generous actions were more rare than relatively selfish actions.

Experiment 1b. Participants (N = 447) from Amazon.com’s Mechanical Turk (MTurk) read about a situation that mirrored Experiment 1a, describing an orchestra that suggests a donation for admission to concerts. The prompt also indicated that “some people donate the suggested amount, some donate more, and some donate less.” We then manipulated the orchestra’s suggested donation to be no specified amount, $10, $20, or $50. Participants then read the same description of Tom W. from Experiment 1a and that he donated $0, $5, $10, $20, or $50 at the concert. Experiment 1b therefore varied not only the target’s behavior but also the expected prosocial conduct.

Salary predictions. Predictions of Tom W.’s salary did not vary across conditions (F < 1). Participants inferred something about Tom W.’s character from his donation rather than about his financial ability to donate.\(^1\)

Experiment 1b. We again created a composite of the warmth items (α = .95) and a composite of the competence items (α = .73). Competence evaluations did not vary across suggested donation amounts, F(3, 446) = 1.69, p = .17, \(\eta_p^2 = .01\). Larger donations led to more positive competence evaluations, F(4, 446) = 8.51, p < .001, \(\eta_p^2 = .07\). The interaction was nonsignificant.

1 Given that participants in this experiment faced the same donation opportunity as the hypothetical Tom W. did, it is possible that participants’ own donation affected their evaluations. Our participants donated $7.93, on average, and 67.0% of participants donated a positive amount. When Tom W. was described as donating $20, most participants experienced an unfavorable social comparison and may therefore have failed to evaluate Tom W. more positively to maintain positive self-regard. However, we observed no significant correlation between participants’ own donation amounts and their evaluations of Tom W.’s warmth in the generous condition, r(23) = −.09, ns, suggesting no contrast effect in that condition. Instead, we observed equally negative correlations in both the fair and selfish conditions, r(31) = −.44, p < .01. This pattern suggests that participants’ own donation was affecting their evaluation of Tom W. but was doing so consistently in the selfish and fair conditions, making it an unlikely explanation for the difference in evaluations between these two conditions.
significant, \( F(12, 446) = 1.29, p = .22, \eta^2_p = .035 \). Again, generosity was not seen as a sign of incompetence.

Evaluations of warmth were of greater interest. A 4 (suggested donation) × 5 (donation amount) between-participants analysis of variance (ANOVA) on warmth yielded main effects for suggested donation, \( F(3, 446) = 9.41, p < .001, \eta^2_p = .06 \), and donation amount, \( F(4, 446) = 81.86, p < .0001, \eta^2_p = .43 \), qualified by the predicted interaction, \( F(12, 446) = 3.85, p < .001, \eta^2_p = .10 \). As shown in Table 2, each donation condition yields asymptotic evaluations, with relatively selfish actions (below the suggested donation) evaluated more negatively as the donation decreases, but no significant effect on evaluations as donations increase beyond the suggested amount.

Discussion

These results suggest asymptotic evaluations of prosociality. It paid to be nice and give the suggested donation, but it did not pay to be nicer and give more. Experiment 1b further demonstrates that the presence of a clearly defined prosocial standard can moderate the asymptotic pattern by determining the point at which evaluations become nonmonotonic. Going above and beyond whatever value was defined as the expected prosocial standard did not create a more positive reputation in the eyes of observers.

At first glance, these results may suggest an experimental artifact: namely, a ceiling effect on warmth evaluations. Although it would still be interesting that evaluations of modestly prosocial actions were of maximum reputational value, we think ceiling effects are an unlikely explanation of our results for two reasons. First, participants’ ratings in the generous condition were not particularly close to the maximum rating of 7, nor were the average evaluations, with relatively selfish actions (below the suggested donation) more negative even directionally more positive as the donation amount increased. Participants could have rated the generous person more favorably but did not. Second, participants’ ratings were not significantly more positive in the fair or generous conditions than in the no information condition of Experiment 1a. There are three possible explanations: A person’s reputation of warmth in the absence of any information is also showing a ceiling effect, participants inferred that Tom donated the suggested amount when it was unspecified, or our participants simply failed to infer a more positive reputation from increased generosity. Of these three, the first seems least likely. However, we address this concern in subsequent experiments by measuring different perspectives on the exchange (Experiment 2), using unbounded measures (Experiment 3), or manipulating the mechanism we believe creates insensitivity to selflessness (Experiment 4b).

Experiment 2: Anticipated and Actual Reputations

Every child learns that sharing candy is the epitome of kindness. We gave one participant (the actor) the opportunity to give candy to another participant (the target) while a third (the observer) watched. By subtle inducement, actors behaved selfishly (gave away only 1 of 10 jelly beans), fairly (gave 5 of 10), or generously (gave 9 of 10). Both observers and targets then reported their impression of the actors’ warmth and competence.

In addition to a conceptual replication, this design can examine another essential aspect of reputational inference: whether people understand how their actions are evaluated by others. We suggested earlier that people may be insensitive to the magnitude of selflessness because they do not spontaneously compare varying levels against each other. Because those who choose courses of action (the actors) are likely to consider several possible outcomes and thus compare between them, we expected they would be more sensitive to their degree of selflessness and would anticipate being valued for their generosity, predicting a more monotonic pattern of evaluations than they actually receive.

Method

Participants (\( N = 195 \)) recruited from a community sample in Chicago completed the experiment for \$2. All participated in groups of three as part of a “tasting experiment.” Participants first completed a self-description questionnaire to provide some individualizing information and then were randomly assigned to role.

Actors received a tray of Jelly Belly’s 10 most popular flavors and were told to choose some for themselves and some to give to

<table>
<thead>
<tr>
<th>Suggested donation</th>
<th>$0</th>
<th>$5</th>
<th>$10</th>
<th>$20</th>
<th>$50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not stated</td>
<td>3.32 (0.78)</td>
<td>5.39 (0.95)</td>
<td>5.37 (0.76)</td>
<td>5.66 (0.58)</td>
<td>5.97 (0.87)</td>
</tr>
<tr>
<td>$10</td>
<td>3.49 (1.18)</td>
<td>4.35 (0.77)</td>
<td>5.38 (0.78)</td>
<td>5.86 (0.65)</td>
<td>5.95 (0.64)</td>
</tr>
<tr>
<td>$20</td>
<td>3.69 (0.85)</td>
<td>4.34 (1.02)</td>
<td>4.96 (0.88)</td>
<td>5.56 (0.87)</td>
<td>5.70 (0.63)</td>
</tr>
<tr>
<td>$50</td>
<td>3.64 (1.44)</td>
<td>4.43 (1.11)</td>
<td>4.79 (0.98)</td>
<td>4.31 (1.05)</td>
<td>5.30 (0.74)</td>
</tr>
</tbody>
</table>

Note. Means that do not share the same subscript across columns differ at a significance of \( p < .05 \). Standard deviations are in parentheses. Cells with italicized contents are those that either meet or exceed the suggested donation amount.

Table 2: Judgments of Warmth as a Function of Suggested and Actual Donations in Experiment 1b
the target. We experimentally manipulated whether actors gave one, five, or nine jelly beans (corresponding to increasingly selfless divisions) by suggesting that it would be a great help to us if they chose a given outcome, but that they were still completely free to choose any outcome they wished (following the standard induced compliance paradigm; Cooper, Zanna, & Taves, 1978).

Actors were informed that targets would not be told how the choice was made. Actors then tasted however many jelly beans they chose from the tray, rated how much they liked each jelly bean (−3 = very bad; 3 = very good), and returned the tray. Actors were then told that the target and the observer would rate them on 10 traits (the five warmth and five competence items). Actors then received targets’ and observers’ self-description questionnaires and predicted how each person would rate them on all 10 items.

Targets were seated in a separate room. They received the tray from the experimenter, who told them that the actor had chosen to give them [one/five/nine] of the jelly beans to taste and ate the rest. Targets were not told of the experimenter’s encouragement to give a particular number, and so the actor’s behavior was presented as a completely free choice. Targets then tasted and rated their jelly beans, as the actors did. Observers sat near the targets and predicted how much the target liked each jelly bean while tasting it. Targets and observers then received the actor’s self-description questionnaire and evaluated his or her warmth and competence.

Results

Choice manipulation. Five actors did not follow the experimenter’s suggestion, instead giving 2, 2, 4, 6, and 8 jelly beans, respectively. Because these participants violated our random assignment, we excluded them from the following analyses (results are slightly stronger if we include them in the condition closest to their choice).2

Liking. As expected, participants liked (or thought another would like) the jelly beans. Liking was significantly positive for each role, ts(59) > 5.59, ps < .001, ds > 1.02. Actors liked their jelly beans (M = 1.96, SD = .88) more than targets did (M = 1.22, SD = 1.38), paired r(59) = 3.65, p < .01, who in turn liked their jelly beans more than observers predicted (M = 0.83, SD = 1.15), paired r(59) = 2.64, p = .01.

Reputations. We created a composite of the warmth items and a composite of the competence items.3 Evaluations of the actors’ competence, as well as the actors’ predicted evaluations, were not affected by the amount of jelly beans given, among either targets or observers (Fs < 2.11, ps > .13).

Again, reputations of warmth were of greater interest. As in Experiments 1a and 1b, observers rated selfish actors as less warm (M = 3.21, SD = 1.11) than fair actors (M = 5.30, SD = .74), t(38) = 7.00, p < .001, d = 2.22, but they did not rate generous actors as warmer (M = 4.16, SD = 1.44) than merely fair actors (see Figure 2). In fact, they rated generous actors as significantly less warm than fair actors, t(38) = 3.15, p = .003, d = 0.99, but still as significantly warmer than selfish actors, t(38) = 2.34, p < .05, d = 0.74. This result is consistent with the inverted-U hypothesis. However, we are reluctant to interpret this result as a general pattern because none of our other experiments replicate this pattern, nor do targets replicate this pattern in this experiment. Given these failures to replicate, we suspect this pattern resulted from an unexpected result among the observers, who thought that targets liked their jelly beans the least. This could make our intended generous act of giving away 9 jelly beans seem significantly less generous for these observers, who might have interpreted it as shirking one’s responsibility in the experiment or as simply giving away mediocre jelly beans.

Targets also rated selfish actors as marginally less warm (M = 4.66, SD = 0.99) than fair actors (M = 5.29, SD = 1.13), t(38) = 1.87, p = .07, d = 0.59, but they did not rate generous actors as any warmer (M = 5.32, SD = 1.14) than fair actors, t(38) = 0.08, p = .93, d = 0.02. These results for both observers and targets again demonstrate that selfishness comes at a reputational cost but that generosity may not produce a reputational benefit beyond equity.

Predicted reputation. Actors, in contrast, expected to be judged in a relatively linear fashion by observers for their prosocial actions, F(2, 59) = 6.43, p < .01, τpb = .16. In particular, actors expected that observers would evaluate them most negatively when they behaved selfishly (M = 4.74, SD = 1.61); nonsignificantly more positively when they behaved fairly (M = 5.17, SD = 1.20), t(38) = 0.96, p = .34, d = 0.30; and even more positively when they behaved generously (M = 6.15, SD = 0.93), t(38) = 2.90, p < .01, d = 0.91. The predictions differed from observers’ actual evaluations, F(2, 57) = 8.72, p < .001, τpb = .23, mainly

![Figure 2. Predicted and actual evaluations of selfish, fair, and generous divisions in Experiment 2. Error bars represent standard errors.](image-url)
because actors did not get the reputational benefit they expected from generosity, paired \( t(19) = 4.73, p < .01 \).

A similar significant, albeit weaker, linear pattern emerged for the actors’ predictions of the targets’ impressions, \( F(2, 59) = 3.16, p = .05, \eta^2_p = .10 \). Actors again expected to be judged most negatively when they behaved selfishly (\( M = 4.64, SD = 1.61 \)); nonsignificantly more positively when they behaved fairly (\( M = 5.34, SD = 1.15 \)), \( t(38) = 1.59, p = .12, d = 0.50 \); and nonsignificantly more positively still when they behaved generously (\( M = 5.77, SD = 1.51 \)), \( t(38) = 1.01, p = .32, d = 0.32 \). These predictions did not, however, differ overall from the targets’ actual evaluations \( (F < 1, \eta^2_p = .02) \).

These results suggest three conclusions. First, they tentatively suggest that people may not completely understand how their good deeds are evaluated by others. Actors expected to be credited to some extent for their generosity by targets and observers, but were not. Second, these results cast further doubt on a ceiling effect limiting positive evaluations of selflessness. In particular, actors expected to be judged significantly more favorably by observers, and directionally by targets, when they behaved generously. If our metrics can detect differences in expected evaluations, then they are also theoretically capable of detecting differences in actual evaluations. Finally, that actors anticipated evaluations were more monotonic than tasters’ actual evaluations suggests that actors were more sensitive to the scope of their prosocial actions than were the targets. This could be because actors had considered other choices they could have made but did not and were therefore more likely to evaluate their chosen option in comparison to other possible alternatives than were targets. Generous actors knew they could have been less generous but chose to follow the experimenter’s suggestion instead. This comparison process for actors may have enabled them to keep their behavior in perspective more than targets and observers did. We test this possible comparison mechanism further in the Experiments 4–6.

**Experiment 3: Unbounded Prosociality**

Our experiments thus far used bounded evaluation measures. These measures leave open the possibility that ceiling effects contributed to our empirical results. We therefore conducted Experiment 3 with different dependent measures, including an unbounded measure of prosociality. In addition, our preceding experiments also tested only three representative actions along the continuum of actions from selfish to generous. Experiment 3 evaluated a larger number of points along this continuum.

**Method**

Mechanical Turk workers (\( N = 182 \)) evaluated Bob D., a person we said came to our laboratory and received $6 to divide between himself and another person. Participants learned that Bob gave away $0, $1, $2, $3, $4, $5, or $6 (manipulated between participants). Participants then reported their impression of Bob’s warmth on three different measures: how much Bob cared for the less fortunate (1 = very little; 7 = very much), the strength of Bob’s selfish motivations (measured with the Selfish Motives scale, an adapted version of the MMPI Cynicism scale; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989)^4, and how much Bob gives annually to charity (estimated in dollars, an unbounded measure of prosociality).

**Results and Discussion**

As shown in Table 3, evaluations were again asymptotic. On every dependent measure, Bob earned an increasingly more negative reputation as he behaved more selfishly (giving $0, $1, or $2) but did not earn an increasingly more positive impression as he behaved more selflessly (giving $4, $5, or $6). These effects also emerged on the unbounded measure, suggesting that ceiling effects do not explain our findings. As Bob made progressively more selfish divisions (giving $2, $1, or $0), participants also predicted he gave progressively less to charity, \( \beta = -.44, t(75) = 4.17, p < .001 \). In contrast, when Bob made progressively more generous divisions (giving $4, $5, or $6), participants did not predict that he gave significantly more to charity, \( \beta = .16, t(78) = 1.39, p = .17 \). Although participants could have evaluated Bob more favorably as he behaved more generously, they did not.

**Experiments 4a and 4b: Single Versus Joint Evaluations**

All preceding experiments suggest an asymmetry in evaluations of relatively selfish versus selfless actions. For one’s reputation, it pays be nice but pays no more to be really nice.

There are two possible explanations for this pattern. One is that the asymmetry reflects objective reputational value, such that selflessness is truly valued according to magnitude whereas selflessness is not. This could be consistent with many different interpretations of Experiments 1–3. Perhaps people value equity so highly in the contexts we have studied that a modestly prosocial action that benefits both the self and others is the ideal solution, and a distaste for inequality mitigates a more favorable impression that would come from an extremely selfless action. A selfish action, in contrast, violates both the preference for equity and prosociality, leading to more monotonic evaluations. Or, perhaps the effects we have demonstrated are consistent with Kant’s (1785/2012) argument that generosity is an “imperfect duty,” meaning that there is no moral imperative for a person to be generous toward others above and beyond fairness. As a result, increasingly generous actions might not be imbued with increasing reputational value. Avoiding extreme selflessness, in contrast, is a “perfect duty” because there is a clear moral imperative to avoid harming others to benefit the self. Increasing selfishness might therefore be judged with increasing moral condemnation. Whatever the particular mechanism, the reputational inferences we have observed may reflect the objective reputational value of prosocial actions.

The second explanation is that people can value increasingly selfless actions monotonically, but their intuitive judgments do not always reflect it. This may occur because people are less likely to evaluate acts of selflessness in context (by spontaneously comparing them against more or less selfish outcomes) than they are acts of selfishness (which may be more readily compared to more or less selfish outcomes). Selfishness is therefore kept in relative perspective, judged in accordance with the magnitude of selfishness.

^4 The MMPI is the Minnesota Multiphasic Personality Inventory. Due to copyright restrictions, we are not able to release the results of this measure here. However, the Selfish Motives measure reveals a pattern consistent with that of the other measures in this experiment, providing further support for our hypotheses. Further details are available by contacting either author.
ness, whereas selflessness is not. As we suggested earlier, this asymmetry in comparison processes could come from the relative rarity of selfless actions compared to selfish actions. Because selfless actions are relatively rare, people lack the requisite familiarity with them to evaluate them according to their magnitude.

These two explanations differ in one essential way. The first suggests that prosociality is not valued monotonically for one of several reasons, whereas the second suggests that prosociality is valued monotonically but that judgments of selfless action may not reflect this value when evaluated in isolation. We test between these two accounts by asking participants to evaluate only one of the possible outcomes (single evaluation mode; Experiment 4a) or by evaluating multiple selfish or selfless outcomes in a fully within-participants design that makes relative comparisons explicit (joint evaluation mode; Experiment 4b). If the asymmetric pattern observed in the preceding experiments reflects objective evaluations of selfless and selfish actions, then we will observe insensitivity to the magnitude of selflessness regardless of the context of evaluation. If the asymmetric pattern reflects a difference in spontaneous comparisons, particularly a failure to judge selfless actions in context, we will observe insensitivity to the magnitude of selflessness when participants evaluate only a single selfless action but will observe sensitivity to magnitude when participants evaluate multiple outcomes.

Method

Experiment 4a. MTurk participants (N = 427) in the selfless condition read about an accomplished professor who received an $80,000 grant for his research and “decided to use $[x$] for his own research and to donate [$80,000−x$] to a nonprofit institution dedicated to research on poverty.” Participants in the selfish condition read about the same accomplished professor, but that he instead found a bag on the street while working containing “a tall stack of $100 bills,” totaling $80,000. Participants then read that he “decided to take $[x$] for himself and to return [$80,000−x$] to the police.” We manipulated how much the professor donated/ took, from $0$ to $80,000, in $10,000 increments (between participants). Note that we do not compare between the selfish and generous conditions, as these scenarios differ on numerous dimensions. We instead test our hypotheses by calculating the slopes of evaluations within each of these conditions.

Because each participant in Experiment 4b had to evaluate nine outcomes, we wanted to streamline the measure of prosociality (evaluating warmth and competence would have required each participant in Experiment 4b to answer 90 questions). Therefore, participants in both experiments simply reported how nice the action was on a scale ranging from 0 (not nice at all) to 100 (very nice).

Experiment 4b. MTurk participants (N = 123) completed the same procedure as 4a, except they evaluated all nine outcomes, from the most generous to the most selfish action, in random order.

Results and Discussion

Experiment 4a. As shown in Figure 3, participants evaluated the giving and taking scenarios differently. Giving nothing appeared less nice than giving $10,000, t(46) = 9.54, p < .001, d = 2.81, whereas taking nothing appeared nicer than taking $10,000, t(46) = 6.10, p < .001, d = 1.80.

Results and Discussion

Experiment 4a. As shown in Figure 3, participants evaluated the giving and taking scenarios differently. Giving nothing appeared less nice than giving $10,000, t(46) = 9.54, p < .001, d = 2.81, whereas taking nothing appeared nicer than taking $10,000, t(46) = 6.10, p < .001, d = 1.80.

Of greater interest were ratings of varying degrees of generosity or selflessness beyond those zero points (between $10,000 and $80,000). Assessing only these outcomes, participants thought that taking more money was increasingly less nice, βstandardized = −.36, p < .001, but did not think that giving more money was increasingly nicer, βstandardized = −.08, p = .28, again reflecting an asymmetric pattern (z = 4.44, p < .001). In fact, giving everything seemed no nicer than giving only one eighth of the money, t(45) = 1.02, p = .31, d = 0.30. Giving money had positive reputational consequences. How much the person gave was irrelevant.

Experiment 4b. Experiment 4b suggests that the asymmetric pattern comes from failing to evaluate selfless actions in relative perspective compared to selfish actions, rather than valuing all selfless actions equally (see Figure 3). In particular, when participants evaluated all possible outcomes, they continued to think that selfishly taking more money was increasingly less nice, βstandardized = −.34, p < .001 (fixed-effects regression). However, they now also valued increasing selflessness, rating more generous actions as significantly nicer than less generous actions, βstandardized = .44, p < .001 (fixed-effects regression). People do value a large act of generosity over a small act, but their judgments of a single act of generosity judged in isolation do not reflect this preference because one act of generosity does not seem to call to

---

Table 3
Evaluations of Varying Degrees of Prosociality in Experiment 3

<table>
<thead>
<tr>
<th>Measure</th>
<th>Amount given</th>
<th>Regression slopes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S0</td>
<td>S1</td>
</tr>
<tr>
<td>Care for others</td>
<td>2.04</td>
<td>2.54</td>
</tr>
<tr>
<td>Gives to charity</td>
<td>$126</td>
<td>$321</td>
</tr>
<tr>
<td>Log of giving</td>
<td>1.69</td>
<td>4.33</td>
</tr>
</tbody>
</table>

Note. Means that do not share the same subscript within each row differ at p < .05. The gives to charity measure was significantly positively skewed (skewness = 2.41, SE = .18; kurtosis = 9.03, SE = .36), and so we present means from both the raw data and a log transformation of the data. The Regression slopes columns report the standardized coefficients across the relatively selfish outcomes (giving $0, $1, and $2) and the relatively selfless outcomes (giving $4, $5, and $6) for each measure, showing that evaluations are sensitive to the degree of selfishness but not to the degree of selflessness. Coefficients marked with an asterisk are significant at p < .05.
that it can pay to be really nice for one’s reputation, but only if observers are reminded that you could have been a little less nice.

**Experiment 5: One Versus 50 Outcomes**

We suggest that reputations are sensitive to the degree of selfishness but not selflessness because the former are kept in context through comparisons whereas the latter are not. One plausible reason for this asymmetry is because selfish actions are simply more common than selfless actions (Engel, 2011) and hence are more likely to bring to mind similar comparison standards that enable more precise evaluations.

In general, the more knowledge or familiarity a person has with any stimulus, the more sensitive the person’s evaluations will be to its magnitude or precise value (Hsee & Zhang, 2010; Morewedge et al., 2009; Parducci, 1965; Stevens, 1975). Evaluations of a student’s ability, for instance, would be more sensitive to that student’s GPA than to the student’s TOEFL score because evaluators are generally more knowledgeable about GPA than about TOEFL scores. If selfish actions are more common and familiar than selfless actions, then selfless actions should be more readily evaluated in relation to their precise magnitude (like GPA) whereas selfless actions should be more insensitive to magnitude (like TOEFL scores). More important, this predicts that making people more familiar with varying degrees of a person’s selfless behavior would enable more monotonic evaluations of prosociality.

We tested this prediction in Experiment 5 by manipulating the amount of information participants had about another person. Instead of evaluating different actions by different people, as in Experiments 4a and 4b, participants evaluated a single person about whom they had little or a lot of knowledge. In particular, participants reported their impression of another person based on a single action or based on 50 actions whose mean outcome was equivalent to the single action. We predicted asymptotic evaluations of prosociality following observation of a single action, as observed in the preceding experiments, but relatively monotonic evaluations based on repeated observations because participants have more comparison information readily available to them.

**Method**

Participants (N = 154; 38% women) were recruited from Amazon.com’s Mechanical Turk and completed the experiment in exchange for a payment equivalent to an hourly rate of $6. We used a 3 (Behavior: selfish, fair, generous) × 2 (Knowledge: single vs. repeated observations) fully between-participants design. In the single observation condition, participants read that “Bob” came to our laboratory to complete a short experiment in which he divided $6 between himself and a stranger (as in Experiment 3). Participants then saw a screen that reported the amount Bob kept for himself and the amount he gave to the other person. Participants saw that Bob gave $1 (selfish), $3 (fair), or $5 (generous) to the other person.

In the repeated observation condition, participants read that Bob came to our lab to complete 50 short experiments in which he divided $6 between himself and a series of strangers. Participants then saw each of Bob’s 50 ostensible choices, shown one at a time on the computer screen, in a random order within each outcome.
condition (see Table 4 for complete details). The average amount given across the 50 choices was $1, $3, or $5, matching the outcome observed in the single-choice condition.

When finished, participants reported their impression of Bob using the same warmth and competence scales used in Experiments 1 and 2. Finally, participants in the 50-choices condition were asked to estimate how many times they saw Bob give $0, $1, $2, $3, $4, $5, and $6.

Results

Reputation. We again averaged the 5 competence ($\alpha = .69$) and 5 warmth items ($\alpha = .95$) together to create one index for evaluations of competence and one for evaluations of warmth.

Competence ratings were not significantly affected by Bob’s behavior, $F(2, 153) = 1.45, p = .24$; the number of observed choices, $F(1, 153) = 2.06, p = .15$; or the interaction between them, $F(2, 153) = 0.59, p = .56$. Generosity was again not seen as a sign of incompetence.

Warmth ratings, however, were again affected by Bob’s prosociality (see Figure 4). A 3 (Behavior: selfish, fair, generous) $\times$ 2 (Observed Choices: 1 vs. 50) ANOVA revealed a main effect for behavior, $F(2, 153) = 111.68, p < .001$, $\eta^2_p = .60$, and a main effect for number of actions, $F(1, 153) = 4.11, p < .05$. $\eta^2_p = .03$, qualified by the predicted interaction, $F(2, 153) = 5.62, p < .01$, $\eta^2_p = .07$. This interaction indicated relatively more monotonic evaluations in the repeated observation condition than in the single observation condition. In the single observation condition, participants evaluated Bob as less warm when he was selfish ($M = 3.34, SD = 1.15$) than when he was fair ($M = 5.98, SD = 0.70$), $t(52) = 9.87, p < .001, d = 2.77$, but they did not evaluate Bob differently when he was generous ($M = 5.78, SD = 0.73$) than when he was merely fair, $t(49) = 1.03, p = .31, d = 0.28$. This replicates the asymptotic pattern observed in evaluations of single actions in all of our experiments thus far.

In contrast, participants in the repeated observation condition evaluated Bob as less warm when he was selfish ($M = 3.33, SD = 1.03$) compared to fair ($M = 4.97, SD = 0.96$), $t(47) = 5.77, p < .001, d = 1.65$, and also evaluated Bob as more warm when he was generous ($M = 5.89, SD = 0.85$) compared to fair, $t(47) = 3.56, p = .001, d = 1.01$. Evaluations were sensitive to the magnitude of generosity only when participants were relative experts: when they had repeated observations about Bob’s behavior to use as a basis of comparison. When participants observed Bob only once, selfishness was evaluated negatively but selflessness was not evaluated significantly more positively than merely being fair.

We made no prediction about whether the asymptotic pattern in the single-evaluation condition is due to people’s overappreciation of small prosocial acts or underappreciation of large prosocial acts. In this experiment, Bob was judged as warmer when participants saw 1 fair action ($M = 5.98, SD = 0.70$) versus 50 actions that on average were fair ($M = 4.97, SD = 0.96$), $t(47) = 4.22, p < .001, d = 1.20$. Evaluations of generous and selfish behaviors did not differ between the single and repeated observation conditions ($ts < .52, ns$), suggesting that participants placed a reputational premium on small acts of kindness judged in isolation rather than undervaluing extremely selfish acts.

Memory. In the 50-choices condition, participants estimated the number of times they saw Bob give each of the 7 giving amounts ($0\text{--}6$). Table 4 presents the relevant means broken down by giving amount. To test whether participants’ memory varied by condition, we averaged the amount participants estimated Bob gave across 50 actions. In the generous condition in which Bob on average gave $5, participants underestimated the average amount Bob actually gave ($M = 4.60, SD = 0.56$), one-sample $t(24) = 3.53, p < .01$. In the fair condition in which Bob on average gave $3, participants’ estimates did not differ from the average amount Bob actually gave ($M = 3.03, SD = .25$), one-sample $t(24) = 0.75, p = .46$. In the selfish condition in which Bob on average gave $1, participants overestimated the average amount Bob actually gave ($M = 1.93, SD = 1.53$), one-sample $t(23) = 2.96, p < .01$. These patterns of recall errors suggest a slight regression to the mean in evaluations but cannot explain the relatively monotonic reputation results in the repeated observation conditions, because participants estimated the generous and selfish conditions to be less extreme than they actually were.

This experiment again demonstrates that people do evaluate generous behavior more favorably than merely fair behavior, as long as they are able to keep the action in perspective by observing multiple actions rather than a single act in isolation. In all experiments reported in this article, observers’ impressions of selfish actions were relatively sensitive to the magnitude of selfishness, whereas impressions of increasingly selfless actions were insensitive to magnitude. One reason selfishness is kept in context, we suggest, is because people are more familiar with varying degrees of selfish actions in everyday life than selfless actions. Providing more information on which to base an evaluation did not alter evaluations of selfish actions, but it created more monotonic evaluations of relatively generous actions. Learning more about a person by watching variance in their behavior over time may create the knowledge base necessary to distinguish between merely nice people and extremely nice people.

In a follow-up experiment reported in the supplementary materials, we provide another test of our evaluability account by providing information about the behavioral norms of a social group rather than the history of a single individual. We find results similar to Experiment 5.

Experiment 6: Friends and Strangers

If keeping generosity in perspective requires having more observations of a person’s behavior for comparison purposes, then judgments of prosociality should be more monotonic when evaluating familiar others than when evaluating unfamiliar others. Friends are more familiar than strangers, and people therefore have more observations of a friend’s behavior than of a stranger’s behavior. Judgments of friends are therefore more likely to be kept in a broader context than judgments of a stranger’s behavior, which is likely to be judged in relative isolation. Our theory about spontaneous comparisons therefore predicts that evaluations of friends’ prosociality would be more monotonic than evaluations of strangers’ prosociality. Experiment 6 tested this prediction.

6 There can be several ways to measure the quality of recall. We also measured recall by calculating participants’ mean deviations from the mean ($MAD$) for each condition, and the results were substantively similar. We believe that method we report is simpler.
Table 4  
Frequencies of Divisions Shown to Participants in the 1-Action and 50-Actions Conditions in Experiment 5

<table>
<thead>
<tr>
<th>Condition</th>
<th>Amount given</th>
<th>Average amount</th>
<th>Single-action amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generous</td>
<td>$0 $1 $2 $3 $4 $5 $6</td>
<td>$5.00</td>
<td>$5.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td> </td>
<td> </td>
</tr>
<tr>
<td>Fair</td>
<td>$0 $1 $2 $3 $4 $5 $6</td>
<td>$3.00</td>
<td>$3.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td> </td>
<td> </td>
</tr>
<tr>
<td>Selfish</td>
<td>$0 $1 $2 $3 $4 $5 $6</td>
<td>$1.00</td>
<td>$1.00</td>
</tr>
<tr>
<td>Recall of these frequencies</td>
<td> </td>
<td> </td>
<td> </td>
</tr>
</tbody>
</table>

Method

Participants (N = 121; 46% women) were recruited in dining halls at the University of Chicago. We used a 2 (Actor: friend vs. stranger) × 3 (Action: selfish, fair, or generous) fully between-participants design. Experimenters approached students at dinner-time and asked if they were interested in participating in an experiment on “the psychology of interactions.” Participants were recruited in pairs, with the stipulation that they had to be “friends for at least a year.” Two pairs of friends participated together, creating four participants per session. We ensured that participants within pairs were friends but that participants across pairs were strangers. This enabled us to create either two pairs of friends or two pairs of strangers in each session.

The experimenter then separated the friend pairs into two separate areas in the dining hall, so that each area contained a pair of strangers. The experimenters ensured that the 2 pairs were always seated at tables that were out of sight of each other. The experimenters explained to participants that they would be randomly assigned to one of two roles, either chooser or receiver. Choosers would be given $2 (real money) and would decide how much of this amount, if any, they would like to share with receivers. In reality, all participants were assigned to be receivers. The experimenters later on brought envelopes to participants, ostensibly from the chooser who were located in the other area in the dining hall. The envelopes contained $0 (selfish action), $1 (fair), or $2 (generous), manipulated between participants.

Experimenters told participants either that their friend was the chooser or that the stranger sitting in the other area of the dining hall was the chooser. Thus, participants thought that they received an amount of money either from their friend or from a stranger.

After seeing the amount of money left for them, participants made two assessments of the chooser. First, they rated how nice the chooser’s action was on a scale ranging from 1 (not at all) to 7 (very nice). Second, they rated the chooser on dimensions of warmth of competence used in Experiments 1, 2, and 5. Of course, participants are likely to have well-formed impressions of their friends’ warmth and competence, impressions that are not likely to be strongly affected by a single action in an experiment. We therefore included these measures to capture likely changes in evaluations of the stranger rather than the friend, where the impression would be based only on the one action observed in the experiment. We did not expect, consistent with basic Bayesian reasoning, that evaluations of a friend’s overall character would be affected by a single action in this experiment.

At the conclusion of the experiment, participants kept the amount of money left for them in the envelopes. They were debriefed, asked not to reveal the details of this experiment to others in the dining hall, and thanked for participating.

Results

Evaluations of actions. We predicted that evaluations of a friend’s action would be relatively monotonic but that evaluations of a stranger’s action would be asymptotic. Table 5 presents the means for each cell in the experiment. A 2 (relationship) × 3 (action) between-participants ANOVA on how nice the action was revealed a main effect for action, F(2, 120) = 125.68, p < .001, η² = .69; no main effect for relationship, F(1, 120) = .06, p > .80; and a marginally significant interaction, F(2, 120) = 2.88, p = .06, η² = .05.

Closer inspection of these results reveals the predicted pattern of evaluations. Among friends, giving $1 was considered nicer (M = 5.09, SD = 1.44) than giving nothing (M = 1.90, SD = 1.07), t(40) = 8.06, p < .001, d = 2.51. More important, giving $2 was considered nicer (M = 6.81, SD = 0.68) than giving only $1, t(41) = 4.95, p < .01, d = 1.53. Evaluations of the friends’ behavior was therefore relatively monotonic.
We did not expect that evaluations of a friend’s general character would be significantly affected by a single action in an experiment. However, evaluations revealed a relatively small but nevertheless monotonic effect. Selfish friends were evaluated as marginally less warm (M = 4.94, SD = 1.72) than fair friends (M = 5.80, SD = 1.36), t(40) = 1.81, p = .078, d = 0.55, and generous friends were evaluated as marginally more warm (M = 6.38, SD = 0.69) than fair friends, t(41) = 1.75, p = .087, d = 0.54.

These findings complement the results of Experiments 4–5 by providing additional evidence that more knowledge about a target enables more monotonic evaluations of prosociality. Being really nice was evaluated more favorably than being merely nice for friends but not for strangers. It pays to be nice, and it can pay even more to be really nice but only when evaluated by familiar others.

General Discussion

Our experiments make the obvious point that it pays to be nice. Participants consistently evaluated those who behaved very selfishly (benefiting themselves but not others) more negatively than those who behaved fairly or equitably (benefiting themselves and others). More important, our experiments also make a nonobvious point: that it may not always pay to be really nice. Participants did not consistently evaluate a person who behaved very selfishly—benefiting others but not the self—more favorably than someone who behaved merely fairly.

This pattern emerged when single events were judged in isolation, suggesting that acts of selfishness were evaluated in context, whereas acts of selflessness were not. Experiments 4–6 make it clear that insensitivity to generosity comes not from failing to value prosociality in others, but rather from failing to compare increasingly generous actions against each other so as to keep them in a broader perspective. Evaluations were more sensitive to increasing degrees of prosociality when actions were compared against each other explicitly (Experiments 4a and 4b), when multiple actions were observed (Experiment 5), or when more was known about a target (Experiment 6). We suggest this asymmetry in evaluations of prosociality stems at least partly from the rarity of selflessness in everyday life compared to selfishness. Commonly observed actions provide a rich knowledge base of similar actions to serve as comparisons, whereas rare actions do not call to mind similar comparisons (Hsee & Zhang, 2010; Kahneman & Miller, 1986; Morewedge et al., 2009; Parducci, 1965; Stevens, 1975). If someone steals a victim’s wallet on the street, even more generous actions are unlikely to spring to mind as sources of comparison. The victim is unlikely to think, “It could have been worse. At least he didn’t steal my car.” But if someone returns a victim’s wallet on the street, even more generous actions are unlikely to trigger similar comparisons.

Beyond these mechanisms, we believe our experiments raise four additional questions: whether these nonlinear patterns of social judgment can explain seemingly nonlinear patterns in social...
behavior, whether these patterns of evaluation are universal or culture bound, why fairness seem to be overvalued in first impressions, and why we failed to identify any negative effects of excessive generosity, as other researchers have observed. We consider each, in turn.

**Prosocial Thoughts Explaining Prosocial Actions?**

We believe the asymmetric pattern of evaluations observed in our experiments may serve as the mechanism of several seemingly unrelated patterns of social behavior. For instance, some research suggests that people are more likely to punish selfishness in others than they are to reward selflessness. In one experiment, observers to a dictator game were given the opportunity to reward or punish dictators (either by giving them additional money or by taking away money; Almenberg et al., 2011). In the most relevant condition of this study, the dictator divided a pot of money between himself and an individual recipient. The observers punished selfish dictators who gave little money to the other person, but they did not reward selfless dictators who gave all of the money away any more than they rewarded the fair dictators who gave away only half of the money. Although this finding was incidental and not discussed in detail in that chapter, this asymmetry is precisely the one we would predict based on the social judgments we observed. If selflessness is not valued more than fairness in social judgment, then it should not be rewarded more than fairness in social behavior, either.

Another recent line of research evaluated reciprocity to selfish, fair, or generous actions (Gray et al., 2014). In these experiments, one person was on the receiving end of a selfish, fair, or generous action and was then given the opportunity to divide a pool of money (as in a dictator game) with a third person. Results consistently demonstrated that people reciprocated selfish and fair actions “in kind,” responding to a third person in a way that roughly matched how they were treated. However, participants did not reciprocate more generously after a generous division than after merely a fair division. Reciprocating equally to fair and generous actions is again what one would expect if the two are valued relatively equally in judgment.

Prosocial judgments guide prosocial behavior, and we would therefore expect the asymmetric pattern we observed here to have a wide variety of behavioral consequences in everyday life. Selfish people may be punished, but extremely generous people may not be socially rewarded any more than only modestly generous people.

**How Universal Are Asymptotic Evaluations of Prosociality?**

Different cultures appear to place different values on prosociality, arguably as a result of complex processes of cultural evolution and current social realities. A recent study examining public goods games in 16 countries revealed that prosociality—behaving selflessly toward others—may actually be punished rather than rewarded in some cultures (Herrmann, Thöni, & Gächter, 2008). Participants in that study had the opportunity to punish others for failing to contribute portions of their endowment for the benefit of everyone else. Whereas participants in all cultures punished extreme selfishness, participants in some cultures also punished extremely generous others.

Because of this finding, it seemed possible that the asymptotic pattern of evaluations we observed may vary across cultures. In particular, those that punish prosociality may actually evaluate it negatively. To test this possibility and to begin examining the universality of our results (Henrich, Heine, & Norenzayan, 2010), we conducted a replication of Experiment 3 with two Russian samples (Klein, Uskul, Grossman, Kraus, & Epley, 2014). In the Herrmann et al. (2008) study, the Russian sample was one of the cultures that punished prosociality the most. We recruited participants (N = 186) in person on a university campus in Novosibirsk and through e-mail solicitations to students at this university. After translating and back-translating the materials from Experiment 3, participants read that “Victor” came to our laboratory and received 180 rubles and was asked to divide that amount between himself and “Nikolai,” a person he had never met before. We then manipulated the amount Victor gave between participants, from completely selfish (0) to completely generous (180), and points in between at 30 ruble increments (30, 60, 90, 120, and 150). These were roughly equivalent to the dollar amounts used in Experiment 3, according to the exchange rate at the time. Participants then evaluated Nikolai on measures of warmth and competence.

A regression revealed an overall positive relationship between warmth and amount given (β = .01, p < .001). The most selfish division (giving 0 rubles) was evaluated more negatively (M = 2.98, SD = 1.08) than the fair division (M = 4.91, SD = 1.06), t(50) = 6.47, p < .001, d = 1.80, but the most generous division (giving 180 rubles; M = 4.84, SD = 1.31) was not evaluated more positively than the fair division, t(46) = 0.20, p = .84, d = 0.06. For generous actions (giving more than 90 rubles), perceptions of warmth did not change as the amount given increased (β = −.003, p = .63). These results replicate the pattern we observed with our American samples and suggest at least two possibilities. First, it could be that the pattern of evaluations we observed is similar across cultures, but the behavior that results from it might vary. Participants may form the same impression of a generous versus a fair person but may, for reasons unknown, nevertheless discourage extreme generosity in others. Second, it could be that we simply failed to replicate the patterns of behavior observed across cultures in the Herrmann et al. (2008) study. Further research is obviously necessary to reconcile these patterns of results. For now, we simply note that our pattern of reputational inferences seems robust, replicating in a culture that some research suggested could yield a very different pattern.

**The Benefits of Fairness**

Understanding how people evaluate prosociality may also shed light on the broad social processes that create a cooperative equilibrium between self-oriented and other-oriented motives. Cooperation is necessary for a flourishing social system (Bowles & Gintis, 2003; Hamilton, 1964; Trivers, 1971), but cooperating with others increases the risk of exploitation and significant personal loss. Considerable amounts of research have studied how violations of cooperation are punished (Barclay, 2004; Fehr & Schmidt, 1999; Feinberg, Willer, Stellar, & Keltner, 2012; Henrich et al., 2010), but much less has studied how social judgments could reward cooperation and kindness in a way that promotes cooperation.
The reputational premium for moderately prosocial behavior we observed in our experiments, coupled with insensitivity to generosity, could be a pattern of evaluations that promotes cooperative exchange without running the risk of exploitation that comes from truly selfless actions. Those who behave fairly quickly develop positive reputations that encourage others to approach them, without suffering the costs of extreme acts of selflessness. These reputational incentives could help explain existing norms for behaving nicely toward others, but not too nicely.

This reasoning is functional in nature and augments the views expressed in developmental and evolutionary explanations of fairness (e.g., Fehr, Bernhard, & Rockenbach, 2008). The reputational premium associated with fair actions (or minimally acceptable prosocial actions) creates an incentive for any individual to behave fairly, because fair actions optimize the ratio of reputational benefit for a given “unit” of personal cost. This may be especially true for zero-sum situations, in which the cost to givers is equivalent to the benefit to recipients, and further research can clarify whether zero-sum and non-zero-sum situations differ in this regard. Broadly, to drive cooperation, reputational incentives may favor not only punishment for selfish actions but rather strong rewards for minimally prosocial behavior. In fact, if behavior is shaped more powerfully by the rewards than by punishments (Bandura, 1971), it could be that the reward given for minimal prosociality is actually a more powerful determinant of future cooperation.

The Trouble With Generosity

The asymptotic pattern of evaluations we observed in isolated judgment is striking because of its failure to value highly generous actions. A neighbor who lends an egg at a time of baking need is no saint; nor is giving 10% of one’s salary to a needy neighbor nearly as nice as giving 100%. In our experiments, participants valued selflessness only when making explicit comparisons between varying degrees of it. This implies that selflessness will be valued simply by observing more of it. We suspect, however, that valuing selflessness is more complicated than this, because there are at least three additional factors that could diminish its apparent value.

First, excessive generosity by others could lead to an unfavorable social comparison with the self and therefore to derogation as a self-defensive strategy. This is especially true when actors voice moralistic motivations for generosity, which may make observers wary of being judged (Monin et al., 2008). Although we did not find reliable evidence of an inverted-U pattern, as this self-defense strategy would predict, our experiments may not have invoked the explicit self-comparisons necessary for others’ generosity to seem threatening.

Second, people strongly believe that self-interest is a powerful determinant of others’ behavior (Epley & Dunning, 2000; Miller & Ratner, 1998). This belief can lead to suspicion that an ulterior motive such as ingratiation is underlying generous actions or to the inference that generous actors do not really value the resources they are giving away. For example, a recent study finds that as people spend more time thinking about the motives for philanthropic behavior they come up with increasingly self-interested attributions (Critcher & Dunning, 2011). We believe this is exactly what happened among observers in Experiment 2, who evaluated generous jelly-bean givers more negatively than merely fair jelly-bean givers. These observers did not seem to credit the givers for their generosity in this case, because they also seemed to believe that the jelly beans were not very desirable. Selfishness seems rarely discounted because it could have been a generous action in disguise. Generosity, in contrast, may be routinely discounted because of cynicism.

Third, the value of generosity may be associated with purity. This may set a particularly high bar for praise, because to be considered generous, actors may be required not only to give generously but also to avoid personal gain from their actions. A recent study finds that prosocial actors may be derogated if they incidentally accrue personal gain from their generous behavior (Lin-Healy & Small, 2013). A blemish on the motivations of generous actors can lead to diminished evaluations, whereas the evaluation of fair actors may remain unchanged even if evidence of their self-interest emerges.

These three reasons, in addition to the comparison processes we documented, may all diminish the reputational value of generosity. Nevertheless, many cultures consider selflessness to be a virtue, as evidenced in the examples of Mother Teresa and Gandhi discussed in the opening paragraph. This suggests a possible disconnect between actual evaluations of prosociality and the rhetoric commonly associated with it. One possible reason for this disconnect is that praise for generosity comes from the experience of self-sacrifice from the actor’s perspective rather than from an observer’s perspective. In Experiment 2, for instance, actors expected that they would be evaluated more favorably for their generosity than they actually were. Another possibility is that the praise for extreme selflessness comes from times in which it is explicitly compared against more moderate acts of kindness. Whatever the cause, those who give of themselves in the hope of being greatly appreciated may end up being sorely disappointed.

References


Received May 22, 2014
Revision received August 22, 2014
Accepted September 1, 2014
Maybe Holier, But Definitely Less Evil, Than You: Bounded Self-Righteousness in Social Judgment

Nadav Klein and Nicholas Epley
University of Chicago

Few biases in human judgment are easier to demonstrate than self-righteousness: the tendency to believe one is more moral than others. Existing research, however, has overlooked an important ambiguity in evaluations of one's own and others' moral behavior that could lead to an overly simplistic characterization of self-righteousness. In particular, moral behavior spans a broad spectrum ranging from doing good to doing bad. Self-righteousness could indicate believing that one is more likely to do good than others, less likely to do bad, or both. Based on cognitive and motivational mechanisms, we predicted an asymmetry in the degree of self-righteousness such that it would be larger when considering unethical actions (doing bad) than when considering ethical actions (doing good). A series of experiments confirmed this prediction. A final experiment suggests that this asymmetry is partly produced by the difference in perspectives that people adopt when evaluating themselves and others (Experiment 8). These results all suggest a bounded sense of self-righteousness. Believing one "less evil than thou" seems more reliable than believing one is "holier than thou."

Keywords: self-evaluation, self-esteem, self-righteousness, social cognition, moral psychology

In matters of morality, many people seem to view themselves through rose-colored glasses. In representative articles, researchers report that "people perceive social reality in ways that support a positive view of themselves" (Goethals, 1986, p. 154), leading "most people [to] think they are more ethical than others" (Fetchenhauer & Dunning, 2006, p. 72) and "chronically feel ‘holier-than-thou’" (Epley & Dunning, 2000, p. 861). Students of psychology routinely read that comparative judgments in moral domains are substantively similar to the more general “better-than-average” effect documented in nonmoral domains: “Compared with people in general, most people see themselves as more ethical, more competent at their job, friendlier, [and] more intelligent” (Myers, 2010, p. 62). These results can be counted on to elicit a knowing chuckle from many audiences, as people seem to readily predict self-serving evaluations in others (Kruger & Gilovich, 1999). Indeed, a widely available bumper sticker echoes the common-sense understanding of the self-concept: “Jesus loves you, but I’m his favorite.” Self-righteousness, at least in Western cultures (Henrich, Heine, & Norenzayan, 2010), seems to know no bounds.

At first glance, this characterization of self-righteousness as unbounded seems based on solid empirical research. Among many other published research results, people report being more likely than others to donate blood, give to charity, give up their seat on a crowded bus for a pregnant woman, treat another person fairly, and generally act prosocially (Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995; Dunning, Meyerowitz, J. & Holzberg, 1989; Epley & Dunning, 2000; Goethals, Messick, & Allison, 1991; Heath, 1999; Heine & Lehman, 1997; Messick, Bloom, Boldizar, & Samuelson, 1985; Miller & Ratner, 1998; Moore & Small, 2007; Paunonen, 1989; Svenson, 1981; Weinstein, 1980). Self-righteousness seems to appear across a wide array of behaviors, whenever people judge their own moral character in relation to others.

We believe, however, that existing research overlooks an important ambiguity in evaluations of moral and ethical behavior, one that may lead to an overly simplistic characterization of people’s well-documented capacity for self-righteousness. In particular, moral behavior includes a broad range of actions that span from doing good to doing bad. Believing that one is more moral than others could reflect a belief that one is either more likely to do good than others (“holier than thou”) or less likely to do bad (“less evil than thou”). When people believe they are more moral than others, do they believe they are more like a saint than others, less like a sinner, or both?

Existing empirical evidence does not clarify this ambiguity because it typically conflates the moral and immoral aspects of behavior. Saying that one would give up one’s seat on a bus for an elderly person, for instance, could be interpreted as a prediction of a moral act of kindness toward another person (making an elderly person comfortable while riding the bus) or as a prediction of avoiding an immoral act that harms another person (not letting an elderly person stand uncomfortably while riding the bus). Understanding the precise nature of self-righteousness requires measuring both characterizations of ethical actions independently.

It is interesting to note that one of the first experiments to document self--other differences in moral judgments (Messick et al., 1985) hints at an asymmetry in self-righteousness. These authors asked people to describe both fair and unfair acts that they have performed themselves and have observed in others. Although not the central component of their analyses, the authors noted an
asymmetry in the severity of the actions people generated. Whereas fair behaviors for self and others were qualitatively similar, unfair behaviors for self and others were noticeably different:

The unfair behaviors that subjects associate with themselves are almost exclusively acts of interpersonal inconsideration. [whereas] the unfair actions associated with others include inconsiderate acts but they also include subcategories that are not found associated with self. Cheating, stealing, destroying, shoplifting, and the like were not found in our sample of “I-unfair” behaviors. (Messick et al., 1985, p. 499)

These results suggest that self-righteousness might be bounded, better characterized by feeling “less evil than thou” rather than “holier than thou.” A similar suggestion comes from research demonstrating that moral behavior is motivated to a surprising degree by people’s discomfort with acting immorally rather than by their desire to act morally (Baumeister, Stillwell, & Heatherton, 1994). People comply with direct requests to help others, for instance, because of discomfort in refusing to help rather than because of a desire to help (Flynn & Lake, 2008).

Despite these preliminary indications, however, research has not explicitly tested the possibility that self-righteousness is asymmetric. In a broad review of the existing literature, we were able to identify only seven articles that offer an incidental test of asymmetric self-righteousness, but all report results in this direction. These results emerge when people list moral and immoral behaviors they and others engaged in (Allison, Messick, & Goethals, 1989; Liebrand, Messick, & Wolters, 1986), recall moral and immoral behaviors they and others engaged in (Gelfand et al., 2002; Green & Sedikides, 2004; Newman, Nibert, & Winer, 2009; Tasimi & Johnson, 2015), and predict the chances that they and others would engage in relatively good and bad behaviors (Dunning & Story, 1991). Although this comparison was not the focus on these experiments, they do suggest an asymmetry in self-righteousness is plausible. Here we report 8 experiments that test this asymmetry hypothesis directly.

Measuring Self-Righteousness, Precisely

There are at least two reasons to predict that self-righteousness will be larger when evaluating immoral behavior than moral behavior. These reasons both stem from the psychological mechanisms that create self–other differences in judgment. First, people tend to evaluate information about others more dispassionately than information about themselves. In particular, information that threatens a person’s identity triggers defensive mechanisms aimed at maintaining a desirable self-image (e.g., Campbell & Sedikides, 1999; Helzer & Dunning, 2012a; Kunda, 1990; Steele & Liu, 1983). People may attempt to discredit the threatening information (Ditto & Lopez, 1992; Ditto et al., 1998), generate new evidence to support a desired belief (Effron, 2014), or even redefine what counts as good evidence to support a favored conclusion (Dawson, Gilovich, & Regan, 2002; Gilovich, 1991). Self–other differences emerge when motivated reasoning “massages” the evidence for one target of evaluation—the self—but not the other. This motivated reasoning is likely to be stronger for threatening stimuli than for flattering stimuli, consistent with a general tendency for negative stimuli to capture attention and guide cognition relatively more than positive stimuli (Baumeister, Bratslavsky, Finkenauger, & Vohs, 2001; Rozin & Royzman, 2001). Imagining oneself donating to charity does not threaten one’s moral self-concept but imagining oneself stealing from a charity does, and so motivated reasoning processes that enable self-righteousness may be triggered more readily for immoral actions than for moral actions (Elliot & Devine, 1994; Festinger & Carlsmith, 1959; Gilbert, Lieberman, Morewedge, & Wilson, 2004).

Second, people naturally adopt different psychological perspectives when evaluating their own versus others’ behavior (Buehler, Griffin, & Ross, 1994; Epley & Dunning, 2000; Kahneman & Tversky, 1979; Koehler & Poon, 2006; Kruger & Gilovich, 2004; Pronin & Kugler, 2007; Williams, Gilovich, & Dunning, 2012). When considering one’s own behavior, people generally adopt an “inside” approach that relies heavily on their knowledge of their own intentions, motives, desires, self-concept, or other person-specific information. People do not, however, have direct access to others’ psychological states and must instead infer them from observed actions. When considering others’ behavior, people therefore tend to adopt an “outside” approach to prediction, basing their predictions on observed behaviors and base rates from which corresponding intentions, motives, and other psychological states are then inferred (Gilbert & Malone, 1995). These differences in perspective can create differences in evaluations of oneself and others when intentions and observed behavior diverge. Because people generally have a positive self-image, people generally construe their own behavior as guided by positive intentions and goals (Markus & Wurf, 1987; Swann & Bosson, 2010; Wilson & Ross, 2001). A boss who lays off employees is likely to construe her behavior as an ethical effort to maintain the company’s profitability for the remaining employees. An employee laid off by the same boss may interpret the boss’s behavior as cruel indifference to the pain of unemployment. As a result, unethical behavior is generally observed in others rather than in oneself (Baumeister, 1999). Notice that this mechanism again suggests an asymmetry in people’s recognition of their own versus others’ capacity for moral versus immoral behavior. Moral behavior seems driven by moral intentions from both an inside as well as an outside perspective, but immoral behavior seems driven by immoral intentions primarily from an outside perspective. This cognitive mechanism therefore predicts more self-righteousness for immoral behaviors than for moral behaviors.

Understanding the precise nature of self-righteousness is important for gaining an accurate understanding of one of the most widely studied research topics in psychology—the self-concept. It is also of practical importance for understanding choices that are informed by one’s self-concept. A person who feels less prone to behaving unethically than others, for instance, might not guard against temptations for unethical behavior. Both the motivational and cognitive mechanisms that create self–other differences in evaluations suggest a bounded sense of self-righteousness. Instead of viewing themselves through rose-colored glasses, people may view themselves through the equivalent of rose-colored bifocals.
Overview of Experiments 1–7

We tested the precise nature of self-righteousness by measuring people’s predictions about their own and others’ moral and immoral behavior (Experiments 1–6), or measuring their memory for past moral and immoral behavior (Experiment 7). In all experiments, we studied participants that existing research suggested would be among the most likely to be self-righteous: Americans (Henrich et al., 2010). If the well-documented tendency among these participants for self-righteousness is unbounded, then we would expect what appear to be self-enhancing judgments for both moral and immoral behavior. If self-righteousness is bounded, then we would expect larger self–other differences when evaluating immoral actions than when evaluating moral actions.1

Experiment 1: Predicting Moral and Immoral Behaviors

Participants predicted whether they were more or less likely than others to engage in each of seven moral and seven immoral behaviors.

Participants also rated how ethical, desirable, and common each behavior was in order to assess the plausibility of several alternative interpretations of our results. Ratings of ethicality and desirability served as manipulation checks of moral and immoral actions. Ratings of commonality tested an alternative interpretation based on egocentrism (Chambers & Windschitl, 2004; Klar & Giladi, 1999; Kruger, 1999). In particular, comparative judgments of the self against others may be based more heavily on assessments of the self than on assessments of others. This egocentrism account predicts that people would believe that they are more likely than others to engage in common behaviors and less likely than others to engage in uncommon behaviors, simply because people can find it hard to imagine themselves engaging in uncommon behaviors but do not think about others being unlikely to engage in uncommon behaviors as well. This egocentrism could provide an alternative account of our predicted results if people think that immoral behaviors are objectively less common than moral behaviors.

Method

Participants (N = 64) were recruited for “a study on people and behavior” from Amazon.com’s Mechanical Turk (M-Turk) and compensated at a rate equivalent to $6 per hour. After reading instructions and answering a practice question, participants read about the seven moral and seven immoral behaviors detailed in Table 1 and answered how likely they are to engage in those behaviors in comparison to other people on a scale ranging from −7 (Others are far more likely to do this than me) to 7 (I am far more likely to do this than others) with 0 as the midpoint (Neither more or less likely for me or for others). For example, the moral behaviors included returning a lost wallet and buying food for a homeless person, and the immoral behaviors included lying to coworkers and stealing small amounts of money (Table 1 contains the exact text). Participants then rated how ethical, desirable (specifically, bad or good), and common the behaviors are on 9-point scales. The order of the 14 behaviors was randomized, such that the seven moral and seven immoral behaviors were interleaved with each other.

Results

As Figure 1 shows, participants believed that they are less likely than others to engage in immoral behaviors, as evidenced by the significantly negative ratings for immoral behaviors (M = −3.50, SD = 2.05), one-sample t(63) = 13.68, p < .001, d = 3.45. Participants did not, however, believe they are significantly more likely than others to engage in moral behaviors (M = .26, SD = 1.78), one-sample t(63) = 1.18, p = .24, d = .30. To compare the magnitude of self-righteousness for immoral and moral behaviors against each other, we first reverse-scored ratings for the immoral behaviors so that larger positive numbers indicate more self-righteousness. As predicted, self-righteousness was larger when evaluating immoral behaviors than when evaluating moral behavior, paired t(63) = 11.89, p < .001, d = 1.50.

Consistent with our intended manipulation, participants rated the seven moral behaviors as more ethical (M = 3.06, SD = .90), on average, than the seven immoral behaviors (M = −2.78, SD = .85), paired t(63) = 31.12, p < .0001, d = 3.88. Notice that these average ratings demonstrate no asymmetry in judgments of ethicality, nor an asymmetry in the extremity of the behaviors we selected. Participants also rated the moral behaviors as more desirable (M = 8.02, SD = .83) than the immoral behaviors (M = 2.16, SD = .87), paired t(63) = 31.06, p < .0001, d = 3.88. This is not surprising given that desirability and ethicality are highly correlated, r(64) = .80, p < .001.

Finally, participants rated the moral behaviors as less common (M = 4.38, SD = 1.12) than the immoral behaviors (M = 4.93, SD = 1.14), paired t(63) = 2.88, p < .01, d = .36. These results suggest that egocentric social comparisons are not producing the observed asymmetry in self-righteousness, because this mechanism suggests that people believe that they would be more likely than others to engage in common behaviors. Instead, our participants believed that they would be substantially less likely than others to engage in the immoral—and to them, common—behaviors. To further test whether commonness ratings affected the asymmetry in self-righteousness, we reverse-scored self–other ratings for immoral behaviors and entered commonness ratings of moral and immoral behaviors as covariates into an analysis of variance (ANOVA). The main effect of moral versus immoral behaviors remained, F(1, 61) = 4.38, p = .041, ηp² = .067.

Experiment 2: Direct and Indirect Comparative Judgments

Comparative judgments can be elicited in two ways: Either by asking participants to directly compare themselves to others on one response scale (as we did in Experiment 1), or by asking participants to indirectly compare themselves to others by first rating their own...
likelihood of engaging in a behavior and then separately rating others’ likelihood of engaging in that behavior. Past research finds stronger self–other differences in direct comparisons than in indirect comparisons (Klar & Giladi, 1997; Moore, 2007; Moore & Kim, 2003). To test whether the asymmetry in self-righteousness depends on elicitation method, Experiment 2 provides a replication of Experiment 1 using both direct and indirect comparison methods.

Method

Participants (N = 108) were master’s of business administration students in a management course who completed this survey as part of a class exercise before the first class meeting. This was a 2 (within-subject factor: moral vs. immoral behaviors) × 2 (between-subjects factor: direct vs. indirect elicitation) design. This survey used the same materials from Experiment 1. Participants evaluated the seven moral and seven immoral behaviors in a random order using either a direct or indirect elicitation method. Participants in the direct comparison condition provided their judgments on a single scale ranging from −7 (Others are far more likely to do this than me) to 7 (I am far more likely to do this than others) with 0 as the midpoint (Neither more or less likely for me or for others). Participants in the indirect comparison condition provided their judgments on two separate scales: One for the self and one for other people, both on scales ranging from 0 (Not at all likely) to 7 (Extremely likely).

| Moral and Immoral Behaviors Used in Experiments 1–2 |
|----------------|----------------|----------------|
| Behavior       | Moral          | Immoral        |
| 1. Stop to help someone with a flat tire.       | 1. Take advantage of a person who does not know the value of a product and sell it to them at an inflated price. |
| 2. Donate blood when asked to do so             | 2. Rush to take the last seat on a crowded bus ahead of an elderly lady. |
| 3. Return a lost wallet you found to the police, leaving the significant amount of cash inside of it untouched. | 3. Find a $20 tip left for the waiter in a restaurant and take the money for yourself. |
| 4. Spend a Sunday volunteering in a soup kitchen. | 4. Engage in an extramarital affair. |
| 5. Tell a professor that he or she had incorrectly marked your final exam and gave you too high a grade. | 5. Lie to your coworkers to increase the chances that you will get a promotion rather than them. |
| 6. Return $20 you had been incorrectly given as change after making a small purchase. | 6. Offer your help in the future while knowing that you do not intend to fulfill the promise when the time comes. |
| 7. Buy food for a homeless person standing outside of a grocery store. | 7. Crash into a parked car and drive off without leaving a note. |

![Table 1](image)
Results

We observe asymmetric self-righteousness in both the direct and indirect methods (see Figure 2). When making direct comparisons, participants predicted they would be significantly less likely than others to engage in immoral behaviors ($M = -1.93$, $SD = 1.71$), one-sample $t(53) = 8.31$, $p < .001$, $d = 2.28$. In contrast, participants did not predict they would be significantly more likely than others to engage in moral behaviors ($M = -0.29$, $SD = 1.31$), one-sample $t(53) = -1.66$, $p = .10$, $d = .46$. As in Experiment 1, the magnitude of self-righteousness was larger when evaluating immoral behaviors than when evaluating moral behaviors, paired $t(53) = 9.66$, $p < .001$, $d = 1.34$.

To create a measure of self-righteousness for indirect comparisons, we subtracted ratings for others from ratings for the self (i.e., [self-ratings] – [other-ratings]). Participants predicted that they would be significantly less likely than others to engage in immoral behaviors ($M = -1.02$, $SD = .75$), one-sample $t(53) = 10.03$, $p < .001$, $d = 2.76$, and also significantly more likely than others to engage in moral behaviors ($M = .41$, $SD = .81$), one-sample $t(53) = 3.70$, $p < .01$, $d = 1.02$. However, the magnitude of self-righteousness was again larger when evaluating immoral behaviors than when evaluating moral behaviors, paired $t(53) = 6.80$, $p < .001$, $d = .92$ (see Figure 2).

Next we compare self-righteousness between the direct and indirect methods. Consistent with past research (e.g., Moore, 2007, Experiment 1), we standardized our measure of self-righteousness across elicitation method because the direct and indirect methods utilize different scales. A 2 (elicitation method: direct vs. indirect) × 2 (behaviors: moral vs. immoral) mixed-model ANOVA revealed a main effect for elicitation method, $F(1, 106) = 42.21$, $p < .01$, $\eta^2_p = .29$, a main effect for behaviors, $F(1, 106) = 60.63$, $p < .01$, $\eta^2_p = .36$, and no interaction, $F(1, 106) = .27$, $p = .61$, $\eta^2_p = .002$. Consistent with prior work, self–other differences were larger in direct comparisons than in indirect comparisons. More important for our hypotheses, self-righteousness was significantly larger when evaluating immoral behaviors than when evaluating moral behaviors. The nonsignificant interaction indicates that the magnitude of asymmetric self-righteousness did not vary by elicitation method.

Experiments 3–4: Predicting Identical Moral and Immoral Behaviors

Experiments 1 and 2 provide initial support for bounded self-righteousness. However, these experiments asked participants to consider qualitatively different moral and immoral behaviors, raising concerns that their results are due to some other confound in these behaviors. Experiments 3–4 therefore provide a more precise test of our hypothesis by providing participants with a fixed set of behaviors, framed as either a moral outcome or an immoral outcome.

Beyond providing a direct replication, Experiment 4 also tests an alternative explanation based on a rather sophisticated social desirability bias in survey responses. That is, participants may exhibit false modesty when asked about the likelihood of engaging in moral behavior, which dampens the self–other difference that characterizes their true self-concept. Participants may actually believe that they are in fact ‘holier than thou,’ but may be unwilling to report this for fear of appearing arrogant. Although our surveys are completely anonymous and confidential, we nevertheless measured socially desirable responding in this replication to assess the plausibility of this alternative interpretation.

Method

Experiment 3. Participants ($N = 79$) were recruited from M-Turk and compensated at a rate equivalent to $6 per hour. Participants read about seven different behaviors in a random order, framed either as moral or immoral. For example, one behavior was crashing into a parked car. The moral frame stated that the driver left a note with his or her contact information. The immoral frame stated that the driver did not leave a note with his or her contact information. Another behavior was getting onto a crowded bus. The moral frame stated that the person gave up the last seat for an elderly lady, whereas the immoral frame stated that the actor rushed to take the last seat ahead of an elderly lady. Table 2 shows full details of all the behaviors and frames.

Participants completed the same measures used in Experiment 1 for each behavior: how likely they were to engage in each behavior compared with others, as well as how ethical, desirable (i.e., good or bad), and common each behavior was. The order of these questions varied randomly within each behavior.

Results and Discussion

As Figure 3 shows, participants predicted that they would be significantly less likely than others to engage in immoral behaviors ($M = -2.31$, $SD = 1.82$), one-sample $t(39) = -8.03$, $p < .0001$, $d = 2.57$, and also significantly more likely than others to engage in moral behaviors ($M = 1.23$, $SD = 2.19$), one-sample $t(38) = 3.51$, $p = .001$, $d = 1.14$. These results again reveal the predicted asymmetry in evaluations because the magnitude of self-righteousness was significantly larger for immoral behavior than for moral behavior, $t(77) = 2.39$, $p = .019$, $d = .54$. The belief that one is “less evil” than others is stronger than the belief that one is “holier” than others.

Consistent with our intended manipulation, participants again rated the moral behaviors as more ethical ($M = 3.03$, $SD = .88$) than the immoral behaviors ($M = -2.21$, $SD = .75$), $t(77) = \ldots$
Notice that the asymmetry observed in behavioral predictions cannot be explained by differences in the extremity of the behaviors we described. Both immoral and moral behaviors differed significantly from the extreme ends of the scale (−4 and +4, respectively), one-sample t(4) > 6.87, ps < .0001, but the moral behaviors were rated as closer to the extreme than the immoral behaviors, z = 2.73, p < .01, suggesting that self-righteousness was smaller for moral behaviors even though the moral behaviors were relatively more ethical than the immoral behaviors were unethical.

Participants also rated the moral behaviors as more desirable (M = 7.97, SD = .82) than the immoral behaviors (M = 2.65, SD = .74), t(77) = 30.20, p < .0001, d = 6.84. Desirability and ethicality were almost perfectly correlated, r(79) = .99, p < .0001.

Finally, participants again rated the moral behaviors as less common (M = 4.52, SD = 1.43) than the immoral behaviors (M = 5.34, SD = 1.08), t(77) = 2.90, p = .005, d = .66. When we reverse-scored ratings of immoral behaviors and entered commonness ratings as a covariate into an ANOVA, the effect of moral versus immoral behaviors remained significant, F(1, 78) = 4.39, p = .040, η²p = .06.

### Experiment 4

These results provide further support for bounded self-righteousness. Using identical behavioral contexts and varying only the ethicality of the outcome, participants again showed more self-righteousness in predictions of immoral behavior than in predictions of moral behavior. As mentioned earlier, one alternative is that these results are produced by a particular kind of social desirability bias, whereby participants exhibit false modesty when asked about the likelihood of engaging in moral behavior. Participants may actually believe that they are ‘holier than thou,’ but may not say this for fear of appearing arrogant. To examine the plausibility of this alternative interpretation, we conducted Experiment 4 as an exact replication of Experiment 3 but added a scale that measures socially desirable responding (the Marlowe–Crowne Social Desirability Scale; Crowne & Marlowe, 1960). This scale asks 33 questions for which answering truthfully requires admitting to small transgressions, such as voting while being ill-informed about the candidates and gossiping about others.

If socially desirable responding is explaining our asymmetry, then the Marlowe–Crowne scale should correlate differentially with immoral versus moral behavior. That is, those who score high on our measure of social desirability might appear especially

---

**Table 2**

**Identical Moral and Immoral Behaviors Used in Experiment 3**

<table>
<thead>
<tr>
<th>Moral behavior [immoral behavior]</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Crash into parked car but leave [without leaving] a note with your contact information.</td>
<td></td>
</tr>
<tr>
<td>2. Give up [rush to take] the last seat on a crowded bus for [ahead of] an elderly lady.</td>
<td></td>
</tr>
<tr>
<td>3. Find a $20 tip left for the waiter in a restaurant and return it to the waiter [take it for yourself].</td>
<td></td>
</tr>
<tr>
<td>4. Tell the truth [lie] to your co-workers even though you know it will increase the chances they will get a promotion rather than you [to increase the chances that you will get the promotion rather than them].</td>
<td></td>
</tr>
<tr>
<td>5. See someone who needs help with a flat tire and stop to help [but ignore them and keep on driving].</td>
<td></td>
</tr>
<tr>
<td>6. Tell [keep] it to yourself when you notice that a professor that he or she [had] incorrectly marked your final exam and gave you too high a grade.</td>
<td></td>
</tr>
<tr>
<td>7. Return [keep] $20 you had been incorrectly given as change after making a small purchase.</td>
<td></td>
</tr>
</tbody>
</table>

---

**Figure 3.** Behavior predictions for self and other for identical behaviors framed as either moral or immoral in Experiment 3 (between-subjects).
self-righteous when considering immoral behaviors and especially modest when considering moral behaviors.

We recruited a larger sample size (n = 188 M-Turk workers) to meet the power criterion needed to detect potential correlations with the Marlowe–Crowne scale. Results again showed evidence consistent with bounded self-righteousness, albeit somewhat weaker than we observed in Experiments 1–2. Participants again predicted that they would be significantly less likely than others to engage in immoral behaviors (M = −2.08, SD = 1.95), one-sample t(92) = −10.29, p < .0001, d = 2.15. Participants also predicted that they would be significantly more likely than others to engage in moral behaviors (M = 1.37, SD = 1.95), one-sample t(94) = 6.83, p < .001, d = 1.41. These results are again consistent with an asymmetry in self-righteousness because its magnitude was significantly larger for immoral behavior than for moral behavior, t(186) = 2.51, p = .013, d = .37.

Second, socially desirable responding does not appear to explain the asymmetry in self-righteousness. To test this, we reverse-scored the immoral behaviors and entered the Marlowe–Crowne ratings as a covariate into an ANOVA. The effect of moral versus immoral behaviors remained significant, F(1, 185) = 6.17, p = .041, n² = .03. Moreover, correlations between behavioral predictions and the Marlowe–Crowne scale do not suggest an asymmetry in socially desirable responding. For immoral behaviors, this correlation was negative, r(93) = −.33, p = .001. The more a participant exhibited socially desirable responding on the Marlowe–Crowne scale, the more this participant believed that he or she would avoid immoral behavior compared with others. For moral behaviors, this correlation was positive, r(95) = .37, p < .001. The more a participant exhibited socially desirable responding on the Marlowe–Crowne scale, the more this participant believed that he or she would enact moral behavior compared with others. Notice that the magnitude of this correlation with the Marlowe–Crowne scale does not differ between immoral and moral behavior, demonstrating no asymmetry in socially desirable responding. If false modesty—a socially desirable response—was guiding people’s reluctance to predict being more moral than others, then we would have observed a negative correlation between the Marlowe–Crowne scale and predictions for moral behaviors. At the very least, we would have observed a different pattern of correlations between the Marlowe–Crowne scale and predictions of immoral versus moral behavior. That we observe correlations of the same magnitude in both cases suggests that false modesty is not a plausible alternative interpretation for bounded self-righteousness.

Experiments 5–6: Framing Moral and Immoral Behavior

Experiments 5–6 provide two additional tests of our hypothesis by examining whether the same objective outcome framed as either moral or immoral would produce an asymmetric pattern of self-righteousness. Participants faced the same hypothetical decision: how to divide money between oneself and a charity. The action underlying the decision, however, was framed as either a relatively moral action (giving money to the charity out of one’s study compensation) or relatively immoral action (taking study compensation from funds earmarked for charity). In Experiment 5, paid online participants evaluated a hypothetical charity in an online survey and predicted how much they would give or take out of $5. In Experiment 6, volunteer participants recruited at a science museum selected a favorite charity out of 3 options and predicted how much they would give or take out of $20. We tested whether self-righteousness—a larger predicted difference between the self and others—was again stronger for an action framed as immoral than for the objectively identical action framed as moral.

Method

Experiment 5. Participants (N = 269) were recruited on M-Turk. We used a 2 (action: give vs. take) × 2 (target: self vs. other) mixed-model design with action varying between-participants and target varying within-participants. After completing an unrelated study (the sample size was set for this unrelated study), participants were asked to imagine that they were given an additional compensation of $5 for the study. In the give condition, participants were asked to imagine that they were also given an envelope that would be sent to a charity organization of their choice if they chose to donate some of their compensation. In the take condition, participants were asked to imagine that their additional compensation was already in the envelope earmarked for charity, but they could take some of this money as additional compensation for themselves (see also Keysar, Converse, Wang, & Epley, 2008; List, 2007). Notice that these choices are objectively identical (how to split $5 between themselves and a charity) but are subjectively distinct (dividing money by giving to a charity vs. taking from a charity). The give condition involves an action perceived to be relatively moral whereas the take condition involves an action perceived to be relatively immoral (Keysar et al., 2008). After each participant predicted how much money he or she would donate or take, all participants were then asked to predict the average amount other people in this experiment would donate to, or take from, a charity organization.

Experiment 6. This procedure was the same as Experiment 5 except that participants were unpaid volunteers recruited from a large science museum (N = 103), participants selected their favorite of three different charities and imagined donating to that specific charity, and participants imagined giving or taking out of $20. We used a 2 (action: give vs. take) × 2 (target: self vs. other) mixed-model design with action varying between-participants and target varying within-participants.

Interested visitors to a large museum in Chicago agreed to complete our survey. Participants were first asked to indicate which of three charities they would be most interested in supporting: the American Red Cross, Society for Prevention of Cruelty to Animals, and the Salvation Army. Participants assigned to the give condition were then asked to imagine that they had participated in an experiment, were paid $20 for their time, and were then given the opportunity to donate some amount of this money to their favored charity. Participants then predicted how much they thought they would donate if they were actually faced with this decision, and also predicted how much the average person would donate. Participants assigned to the take condition, in contrast, were asked to imagine that they participated in an experiment, and that $20 would be given to their preferred charity in exchange for their time, but that participants could keep some of this money for themselves. These participants then predicted how much of this $20 they
would take for themselves, and how much the average person would take. Participants then answered some demographic questions and were debriefed.

Results and Discussion

Experiment 5. As Table 3 shows, participants in the take condition predicted that they would take significantly less money from the charity ($M = 2.78, SD = 1.93) than other people would ($M = 3.35, SD = 1.38), paired $t(133) = 3.66, p < .001, d = .33$. Participants in the give condition, however, did not predict that they would give significantly more money to charity ($M = 1.85, SD = 1.68) than other people would ($M = 1.91, SD = 1.30), paired $t(134) = .29, p = .77, d = .03$. A 2 (action: give vs. take) × 2 (target: self vs. other) ANOVA with repeated measures on the second factor revealed a main effect for target, $F(1, 267) = 6.64, p = .01, 
\eta^2_p = .02$, a main effect for action, $F(1, 267) = 6.68, p < .01, \eta^2_p = .01$, qualified by a predicted interaction, $F(1, 267) = 8.73, p = .003, \eta^2_p = .03$. Self-righteousness, defined as the predicted difference in behavior between oneself and others, was again asymmetric.

Experiment 6. As in Experiment 5, participants in the take condition predicted that they would take significantly less money from the charity ($M = 2.29, SD = 5.50) than others would ($M = 6.73, SD = 6.17), paired $t(50) = 4.69, p < .001, d = .66$. Unlike Experiment 5, participants in the give condition also predicted that they would give significantly more money to charity ($M = 10.94, SD = 11.48) than others would ($M = 7.38, SD = 10.34), paired $t(51) = 2.82, p = .007, d = .40$. Overall, these volunteer participants recruited from a Museum predicted being much more generous than did the online participants recruited for pay in Experiment 5. Indeed, 80.4% of participants in the take condition predicted they would take nothing from the charity out of $20$, whereas only 20.9% of participants in Experiment 5 said they would take nothing. In order to test for an asymmetry in self-righteousness in a data set with such massive floor effects in one cell of the design, we conducted a Tobit regression. This test is specifically designed for analyzing censored data where parametric tests are inappropriate. We first reverse-scored responses in the take condition so that higher numbers indicated more self-righteousness. As predicted, a Tobit regression with participant fixed-effects revealed a significant action (take vs. give) by actor (self vs. other) interaction consistent with asymmetric self-righteousness, $\beta = -7.52, SE = 3.17, z = 2.37, p = .018$.

Experiments 5 and 6 both provide evidence consistent with bounded self-righteousness. It is interesting to note that participants in both experiments believed they would take significantly less money from a charity than others would, but were less certain that they would give significantly more money to a charity than others would. That participants in Experiment 6 showed evidence of self-righteousness when predicting moral behavior but participants in Experiment 5 did not likely stems partly from objective differences between our samples. Participants in Experiment 5 were recruited for pay in an online labor market (for some a regular job) and were imagining a hypothetical charity, whereas participants in Experiment 6 were likely to be more financially stable, were volunteering their time, and were imagining a specific charity they just chose as one of their favorites. The degree to which people think they are more generous than others depends on many different factors, but the asymmetry in evaluations of moral versus immoral behavior seems consistent across them.

Experiment 7: Remembering, Self-righteously

Experiment 7 provides one final independent test for the existence of bounded self-righteousness by examining whether an asymmetry exists in memory for immoral versus moral behavior. Because remembering the past is at least partly a process of construction that relies on the same psychological mechanisms as predicting the future (Schacter, Addis, & Buckner, 2007), the basic motivational and cognitive mechanisms that create bounded self-righteousness in predictions of future behavior in Experiments 1–6 should also create bounded self-righteousness in memory for one’s own and others’ behavior. Following a procedure similar to Messick et al. (1985), we asked participants to recall recent immoral and moral actions committed by themselves and others. We expected that participants would recall more extreme immoral actions for others than for themselves, but would not recall systematically more moral actions for themselves than for others.

Method

Participants ($N = 247$) were recruited from M-Turk. We used a 2 (behavior frame: ethicality vs. altruism) × 2 (target: self vs. other) mixed-model design, with behavior frame and target varying between-subjects and action varying within-subjects. To examine whether the results observed in previous experiments might be produced partly by the way we framed moral and immoral behavior, we manipulated this description in this experiment. In the ethicality condition, participants read that the experiment was investigating “the psychology of ethical and unethical behavior.” In contrast, in the altruism condition, participants read that the experiment was investigating “the psychology of generous and selfish behavior.” Participants then

Table 3

<table>
<thead>
<tr>
<th>Exp.</th>
<th>Nature of judgment</th>
<th>Moral condition</th>
<th>Immoral condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Self</td>
<td>Other</td>
</tr>
<tr>
<td>5</td>
<td>Predict giving to or taking from charity ($5$)</td>
<td>$1.85_a$</td>
<td>$1.81_a$</td>
</tr>
<tr>
<td>6</td>
<td>Predict giving to or taking from charity ($20$)</td>
<td>$10.94_a$</td>
<td>$7.98_b$</td>
</tr>
<tr>
<td>7</td>
<td>Recall moral or immoral action</td>
<td>$2.96_a$</td>
<td>$3.14_a$</td>
</tr>
</tbody>
</table>

Note. The “Difference” columns represent absolute differences between means [Self – Other]. Different subscripts within rows denote means that differ at $p < .05$. Asterisks in the “Difference” columns denote significant differences from 0 (implying no self-righteousness) at $p < .05$. Exp. = experiment.
read that they would be asked to recall one ethical (or generous) action and one unethical (or selfish) action, according to condition, in counterbalanced order. The framing of participants’ behavior was therefore manipulated between-participants, and the action (moral vs. immoral) was manipulated within-participants.

In a subsequent screen, participants in the self condition read that they should recall actions they engaged in recently. Participants in the other condition read that they should recall actions that they “observed someone you know do recently.” The target was therefore manipulated between-participants. Next, participants wrote about the actions they recalled. Participants were encouraged to “write enough so that we would be able to understand the action itself and the situation and circumstances surrounding it.”

We later asked two independent coders, blind to our hypotheses and our full set of experimental conditions, to evaluate participants’ actions. We sought a measure of positivity—whether an action was positive or negative—to test whether self–other differences emerged on moral behaviors, immoral behaviors, or both. We therefore asked the two coders to rate how extreme each action was on a scale ranging from 7 (extremely positive) and –7 (extremely negative) with 0 as the midpoint (neutral).

### Results and Discussion

We averaged the coders’ evaluations, $r = .91, p < .001$, into a single composite for each action. A 2 (behavior type: ethicality vs. altruism) × 2 (action: moral vs. immoral) × 2 (target: self vs. other) ANOVA on behavior positivity with repeated measures on the second factor revealed main effects for action, $F(1, 243) = 1.837.41, p < .0001$, $\eta_p^2 = .08$, for target, $F(1, 243) = 14.76, p < .001$, $\eta_p^2 = .06$, and for behavior type, $F(1, 243) = 8.94, p < .001$, $\eta_p^2 = .035$. Moral actions were rated more positively than immoral actions, participants recalled more positive behaviors for themselves than for others, and altruistic behaviors were rated more positively than ethical behaviors. More important, a predicted interaction occurred, $F(1, 243) = 26.24, p < .001$, $\eta_p^2 = .097$. As can be seen in Table 3, participants’ own immoral actions ($M = -1.83, SD = 1.53$) were evaluated as less negative than others’ immoral actions ($M = -2.91, SD = 1.71$), $F(1, 245) = 27.24, p < .001$, $\eta_p^2 = .10$. In contrast, participants’ own moral actions ($M = 2.96, SD = 1.25$) were not evaluated differently from others’ moral actions ($M = 3.14, SD = 1.25$), $F(1, 245) = 1.34, p = .25$, $\eta_p^2 = .005$.

Peopleremembered others being more immoral in the past than they were themselves, but did not recall themselves being systematically more moral in the past than others. Self-righteousness appears bounded not only in predictions of the future but also in memory of the past.

### Experiment 8: Explaining Bounded Self-righteousness

Experiments 1–7 provide convergent evidence that self-righteousness is bounded. Although participants thought they were less likely than others to engage in relatively immoral behaviors, they did not think they were more likely than others to engage in relatively moral behaviors. Experiment 8 was designed to test one plausible mechanism underlying this result.

As we explained earlier, existing research suggests two mechanisms that led us to expect this pattern of bounded self-righteousness. One relies on a motivation to think well of oneself. Anything that threatens this self-concept, such as the thought of engaging in immoral behavior, could trigger motivated reasoning that maintains a desirable self-view (Ditto & Lopez, 1992; Steele & Liu, 1983). The thought of engaging in moral behavior does not threaten one’s self-concept as an ethical person, and so would not be expected to produce the same degree of self-righteousness.

Here, however, we test the second well-known mechanism underlying self-righteousness, namely that people assess themselves and others from different perspectives that rely on different sources of information. When evaluating themselves, people rely on “inside” information as part of their self-concept, such as their own intentions, motives, and desires (Buehler et al., 1994; Epley & Dunning, 2000; Kahneman & Tversky, 1979; Koehler & Poon, 2006; Kruger & Gilovich, 2004; Pronin & Kugler, 2007; Williams et al., 2012). Because people generally understand their own behavior in a way that is consistent with a positive self-image, people rarely perceive their own actions as driven by malicious intentions or misanthropic motives (Baumeister, 1999; Swann & Bosson, 2010). As a result, people would be relatively unlikely to predict that they would engage in immoral behaviors. When evaluating others, in contrast, people rely on more easily accessible “outside” information, such as population base-rates and observations of others’ behaviors (Epley & Dunning, 2000; Gilbert & Malone, 1995; Pronin, 2009). Because people observe others engage in both moral and immoral behaviors, relying on “outside” information to predict others’ behaviors will lead to more symmetric predictions about others’ propensity for moral and immoral behaviors. This mechanism therefore creates a larger self–other gap in evaluations of relatively immoral actions than in evaluations of moral actions.

Experiment 8 examines this proposed mechanism by testing whether providing individuating “inside” information about others attenuates self-righteousness in predictions of immoral actions (Experiment 5b, Epley & Dunning, 2000). We asked an initial group of 100 participants to write about their character and “who you are” as a person. We then showed these descriptions to a second group of participants. We predicted that after gaining access to others’ self-evaluations, people would incorporate this information into their assessments of others’ possible behaviors. Because others almost invariably think of themselves as ethical, we predicted that providing this inside information would attenuate the self-righteousness we observed in predictions of immoral behavior.

### Method

#### Stimuli development.

We first recruited 100 participants (52% women) from M-Turk. These participants were asked to write five words that describe them and a short essay describing

---

2 An uninteresting interaction between action and behavior type also emerged, $F(1, 243) = 52.33, p < .001$, $\eta_p^2 = .18$. Immoral actions in the altruism domain were judged to be more negative ($M = -2.99, SD = 1.78$) compared with the altruism domain, ($M = -1.33, SD = 1.39$), $t(245) = 6.16, p < .001, d = .78$. In contrast, moral actions in the altruism domain ($M = 3.33, SD = 1.40$) were judged to be more positive compared with the altruism domain ($M = 2.77, SD = 1.02$), $t(245) = 3.61, p < .001$,$ d = .46$. No other interactions emerged, $Fs < .15, ps > .28$. 

---
“who you are.” We used these responses as individuating information for our main experiment.

For replication purposes, these participants also made comparative assessments about the relative likelihood that they and other people will engage in seven moral and seven immoral behaviors, using the same stimuli used in Experiment 1 (displayed in Table 1). These judgments were made on scales ranging from −7 (Others are far more likely to do this than me) to 7 (I am far more likely to do this than others) with 0 as the midpoint (Neither more or less likely for me or for others). The results indeed replicated the bounded self-righteousness results found in previous experiments. Participants believed that they were less likely to engage in immoral behaviors than others, as evidenced by the significantly negative ratings given to immoral behaviors (M = −3.50, SD = 2.48), one-sample t(99) = −14.11, p < .0001, d = 2.84. But participants did not believe they are more likely than others to engage in relatively moral behaviors (M = .29, SD = 2.34), one-sample t(99) = 1.23, p = .22, d = .25. As in Experiment 1, the magnitude of self-righteousness was significantly larger when evaluating immoral behaviors than when evaluating moral behavior, paired t(99) = 11.25, p < .001, d = 1.13. Bounded self-righteousness appears robust in this population.

Main experiment. Participants (N = 202; 52.5% women) were recruited from M-Turk. We used a 2 (individuating information: yes vs. no) × 2 (behaviors: moral vs. immoral) mixed-model design, with the first factor manipulated between-participants and the second manipulated within-participants. Participants read that they would be asked to assess their own and another randomly chosen person’s behavior. In the individuating information condition, we provided participants with individuating information about one other person, using the stimuli we collected from the 100 participants run in the preliminary phase. In the no individuating information condition, participants did not see a description collected from the preliminary phase of this experiment and were simply told that the other person they will be evaluating is randomly chosen from the sample of M-Turkers participating in this experiment.

All participants then read the seven moral and seven immoral behaviors used in Experiment 1 (randomly ordered) and answered whether they or the other person is more likely to engage in those behaviors on a scale ranging from −7 (This other person is far more likely to do this than me) to 7 (I am far more likely to do this than others) with 0 as the midpoint (Neither more or less likely for me or for others). As in Experiments 1–2, participants also rated how ethical and common the behaviors are on 9-point scales. The order of moral and immoral behaviors was randomized, as was the order of individual behaviors within the moral and immoral categories.

Results and Discussion

Manipulation checks. We first tested whether the moral behaviors were rated as more ethical than the immoral behaviors. A 2 (individuating information: yes vs. no) × 2 (behaviors: moral vs. immoral) ANOVA on ethnicity with repeated measures on the second factor revealed only a main effect for behaviors, F(1, 200) = 1510.95, p < .0001, $\eta_p^2 = .88$. Participants rated the moral behaviors as more ethical (M = 2.83, SD = 1.01) than the immoral behaviors (M = −2.40, SD = 1.21). There was no effect of individuation information on judgments of ethnicity, either for ethical or for unethical behaviors, ts < 1.28, ps > .20.

As in Experiment 1, the immoral and moral actions did not differ in their extremity. Both immoral and moral behaviors significantly differed from the low and high ends of the scale, respectively, one-sample ts > 16.47, ps < .0001, ds > 2.32. However, these differences-from-the-extremes did not differ from each other, z = .97, p = .33. Also consistent with Experiment 1, participants rated the moral behaviors as less common (M = 4.58, SD = 1.12) than the immoral behaviors (M = 5.26, SD = 1.20), paired t(201) = 5.95, p < .001, d = .42. Egocentric social comparisons do not appear to explain bounded self-righteousness.

Comparative judgments. A 2 (individuating information: yes vs. no) × 2 (behaviors: moral vs. immoral) ANOVA on comparative judgments with repeated measures on the second factor revealed a main effect for behaviors, F(1, 200) = 43.13, p < .001, $\eta_p^2 = .18$, a main effect for individuating information, F(1, 200) = 5.25, p = .023, $\eta_p^2 = .026$, and an interaction, F(1, 200) = 10.76, p = .001, $\eta_p^2 = .051$. As Figure 4 shows, the individuating information did not alter the degree of self-righteousness in evaluations of moral behaviors. For moral behaviors, having access to others’ self-evaluations did not affect self–other judgments (M = −.08, SD = 2.16) compared with not having this individuating information (M = 2.23, SD = 2.06), F(1, 200) = 1.14, p = .29, $\eta_p^2 = .006$. Self-righteousness was nonsignificant in both of these conditions, one-sample ts < 1.16, ps > .25, ds < .23. In contrast, having access to others’ self-evaluations altered the degree of self-righteousness for immoral behaviors. Participants who had individuating information reported a smaller self–other difference in the likelihood of engaging in unethical behaviors (M = −.87, SD = 2.31) than participants who did not have individuating information (M = −2.11, SD = 2.36), F(1, 200) = 14.39, p < .001, $\eta_p^2 = .067$. Although participants in both conditions felt they were significantly less likely than others to behave immorally, this difference was more than two times larger when participants lacked individuating information, one-sample t(102) = −9.11, p < .0001, d = 1.80, than when they had it, one-sample t(98) = −3.73, p < .001, d = .75. Comparing these two conditions directly against each other indicates that providing individuating information about the other person’s self-evaluation significantly reduced the magnitude of self-righteousness, t(200) = 3.79, p < .001, d = .54.

![Figure 4. Self–other judgments as a function of morality and individuation in Experiment 8.](image)
The different perspective people adopt when evaluating themselves versus a random “other” does not explain all of the bounded self-righteousness we observed, but it does seem to explain a significant amount of it. Diminishing this perspective gap by providing others’ self-evaluations, the same kind of information people are likely to rely on when predicting their own behavior, significantly reduced the tendency to feel “less evil” than others.

**General Discussion**

It is often said that people view themselves through rose-colored glasses, but our research on self-righteousness suggests something closer to rose-colored bifocals. In a series of 8 experiments, we find convergent evidence for a bounded sense of self-righteousness. Consistent with a large body of existing literature, our participants consistently believed they would behave more ethically than others, but this was primarily true when considering relatively immoral actions. When considering relatively moral actions that involve doing good rather than doing bad, we found relatively little evidence that people thought they would behave more ethically than others. Our 8 experiments contain 11 independent comparisons between self-righteousness for immoral versus moral behaviors (see Figure 5). The average self-righteousness effect for immoral behavior was very large ($d = 1.84$), and significantly greater than zero, one-sample $t(10) = 5.70, p < .001$. In contrast, the average self-righteousness effect for moral behavior across these 10 comparisons was relatively smaller ($d = .45$), albeit still larger than zero, one-sample $t(10) = 2.93, p = .015$. These results did not seem to be produced by a sophisticated pattern of social desirability biases (such as false modesty for moral behavior), as socially desirable responding in Experiment 4 was positively correlated with self-righteousness for both moral and immoral actions. Self-righteousness, at least among the populations we sampled from, may be better characterized as feeling “less evil than thou” than feeling “holier than thou.”

We predicted this bounded sense of self-righteousness based on both the motivational and cognitive mechanisms that are known to produce self–other differences in judgment (Kunda, 1990; Chambers & Windschitl, 2004). Anything that threatens a person’s identity can trigger motivated reasoning to support an existing belief about oneself. Because most people believe they are moral, imagining oneself committing an immoral action could trigger reasoning with the goal of defending an existing positive self-concept (such as by derogating others; Fein & Spencer, 1997). We did not, however, test this motivated reasoning mechanism directly in our research. Instead, we tested a cognitive mechanism based on the difference in perspectives people adopt when evaluating themselves versus others. When evaluating oneself, people tend to adopt an inside perspective, predicting their behavior based on their own intentions, aspirations, or self-concept (e.g., Buehler et al., 1994; Epley & Dunning, 2000; Helzer & Dunning, 2012b; Koehler & Poon, 2006; Kruger & Gilovich, 2004). Because most people think of themselves as ethical, and as having ethical intentions, people are unlikely to believe they would behave unethically. When evaluating others, in contrast, people are more likely to adopt an outside perspective, basing their predictions on observed base rates of behavior in daily life (Buehler et al., 1994; Epley & Dunning, 2000). Because people do indeed learn about others behaving both ethically and unethically, these different perspectives for oneself and others predict a larger self–other difference in predictions of unethical behavior (“I’d never pass by someone with a flat tire along the highway without helping, but I see many others driving by...”)

![Figure 5](image-url)
without stopping, so I must be more likely than others to stop and help someone fix a flat tire"). Consistent with this account, providing people with others’ self-concepts in Experiment 8 significantly diminished self-righteousness.

These results join a growing body of research providing a more precise understanding of self-evaluations. A person’s self-concept is produced by a mix of psychological processes, some of which lead to unrealistically positive self-views but also others that do not. For instance, people tend to evaluate themselves compared with others egocentrically, focusing on their own traits and abilities and only subsequently considering others’ traits and abilities. This can lead to unrealistically positive self-evaluations when people are evaluating themselves compared with others on relatively easy tasks in which everyone is relatively proficient (such as operating a computer mouse), but can lead to unrealistically negative evaluations when people are evaluating relatively difficult tasks (such as juggling a computer mouse; Kruger, 1999; Klar & Giladi, 1999). People may likewise be egocentric when considering the likelihood that certain events will happen to themselves versus others, leading to what appear to be unrealistically optimistic self-views when evaluating relatively common events that are likely to happen to almost everyone but unrealistically pessimistic self-views when evaluating relatively uncommon events that are unlikely to happen to anyone (Weinstein, 1980; Kruger & Burrell, 2004). Finally, people tend to define traits, such as “leadership,” egocentrically by focusing on the traits they possess rather than on the traits they lack, leading to unrealistically positive self-evaluations when considering relatively ambiguous traits (e.g., leadership) but not unrealistically biased evaluations of more concrete traits (such as intelligence; Dunning et al., 1989).

Our research likewise identifies an important boundary on unrealistically positive self-evaluations. Moral behavior spans a broad spectrum of behaviors, ranging from doing bad to doing good. Looking at only part of that spectrum, or conflating the two within a single evaluation, produces an imprecise understanding of how people think about themselves compared with others. Although few are strangers to self-righteousness in everyday life, it may not be as widespread across the moral spectrum as existing research might imply.

Our experiments test how widespread self-righteousness might be across the moral spectrum, but they do not test how widespread this pattern might be across the globe. We have referred to “people” without qualification throughout this article, because we have been referring to those people who are known to reliably exhibit self-righteousness (namely, Westerners from an individualistic culture). Variance in the mechanisms that produce self-righteousness could moderate the results we have observed in other cultures. In particular, some research suggests that self-enhancement motives may be diminished or absent entirely in more collectivist cultures (Heine, 1999; Heine & Kitayama, 1999; Heine, 2005), whereas other research suggests that self-enhancement motives are universal but that their manifestations vary by cultural contexts (Sedikides et al., 2003; Sedikides et al., 2005). In the most direct test of cross-cultural differences in self-righteousness that we know of (Balcetis et al., 2008), the results were somewhat mixed. In this series of four reported experiments, two showed no significant cross-cultural differences in self-righteousness as measured by behavioral predictions, one showed significantly weaker self-righteousness among collectivist participants, and one found directionally weaker self-righteousness among collectivist participants but did not report whether the difference was statistically reliable. A more precise understanding of the nature of self-righteousness may also provide a more precise understanding of cross-cultural differences that may or may not emerge in direct experimental tests. Larger cross-cultural differences may emerge, for instance, in evaluations of unethical behavior than in evaluations of ethical behavior. Whether bounded self-righteousness is moderated by major cultural differences, and whether it can explain variance in the existence of cross-cultural differences, are very important questions for future research.

In the meantime, we believe a more precise understanding of self-righteousness is important because it suggests a specific type of misunderstanding about oneself. In particular, participants in our experiments consistently believed that they would not be as prone to unethical behavior as others. When put to the actual test, however, these predictions about others’ behavior seem to be more accurate than people’s predictions about themselves (e.g., Epley & Dunning, 2000). In one experiment (Kawakami et al., 2009), White participants predicted that they would be less likely to work with another White participant who made a blatantly racist joke than people who heard the joke actually were. In another (Bocchiaro et al., 2012), less than 4% of participants predicted that they would be willing to obey an experimenter’s request to recommend a blatantly unethical and harmful experiment to a potential group of participants. When faced with the actual unethical request, 77% agreed to recommend the unethical experiment.

A potentially mistaken belief about the likelihood of one’s own unethical behavior is precisely what makes the classic experiments of social psychology—such as Milgram (1963), Darley and Latane (1968), and Zimbardo (2007)—so surprising. Participants in these classic experiments behave less ethically than readers of the experiments expect they would behave themselves (e.g., Bierbrauer, 1979). A mistaken sense of self-righteousness may make people in their professional or personal lives unlikely to take steps that would mitigate the risk of unethical actions because they underestimate the likelihood of falling prey to them. A scientist who doesn’t believe she is as likely as others to manipulate data might not adopt laboratory practices that would diminish the temptation (Simmons et al., 2011). A doctor who doesn’t believe he would recommend a drug to a patient simply because he was paid by the drug company might reject conflict-of-interest reforms (Sharek et al., 2012). A gun owner who never believes he would turn his gun on a family member might store it loaded in his nightstand rather than unloaded in a basement safe. People’s self-evaluations matter because they guide their choices. (Swann & Bosson, 2010). A mistaken belief about the likelihood of one’s own ethical risks compared with others might lead people to put themselves in risky ethical situations that they would otherwise avoid.

Concluding Thought

Any statistician knows that paying attention to only a small range of available observations can lead to mistaken inferences. An educator interested in whether SAT scores predict college GPA might see no relationship when looking only at students admitted
to a prestigious university, but would see a very strong correlation when looking across the entire range of SAT scores. A pollster who surveys only elderly citizens might predict a very different outcome in an election than one who surveys the citizens from the entire age spectrum. And a psychologist interested in how people make choices by studying how people respond only to potential losses would miss the very different pattern of behavior observed when examining how people respond to gains (Tversky & Kahneman, 1974). Our research suggests that existing psychological research could paint a misleading picture of the precise nature of self-righteousness, at least partly because it has not systematically examined the entire spectrum of moral actions. People’s tendency toward self-righteousness has been recognized long before researchers documented its magnitude. But what casual observers of others’ behavior can miss are the limits of self-righteousness that empirical research can detect. Examining the entire range of ethical actions reveals the boundaries on self-righteousness that a narrower focus might conceal.

References


References


Received September 17, 2015
Revision received February 23, 2016
Accepted February 23, 2016

Correction to Wilcox et al. (2016)

In the article “How Being Busy Can Increase Motivation and Reduce Task Completion Time” by Keith Wilcox, Juliano Laran, Andrew T. Stephen, and Peter P. Zubcsek (Journal of Personality and Social Psychology, 2016, Vol. 110, No. 3, pp. 371–384. http://dx.doi.org/10.1037/pspa0000045), the affiliation of the author Andrew T. Stephen was incorrectly listed in the byline and the author note. The author is affiliated with the University of Oxford. The author note paragraph “Andrew T. Stephen is now at the University of Oxford” should have been omitted. All versions of this article have been corrected.

http://dx.doi.org/10.1037/pspa0000054