**Social Perception in the Volunteer's Dilemma:**

**Role of Choice, Outcome, and Expectation**

Patrick R. Heck

Joachim I. Krueger

Brown University

8,305 Words

Author Note

Correspondence:

Patrick R. Heck or Joachim I. Krueger
Department of Cognitive, Linguistic & Psychological Sciences
Brown University
190 Thayer St.
Providence, RI 02912
Phone: (401) 863-2503
E-mails: Pheck1000@gmail.com; Joachim@brown.eduP
ages: Heck: www.patrickrheck.com
          Krueger: http://research.brown.edu/research/profile.php?id=10378

**Abstract**

The volunteer's dilemma (VoD) is a challenging interpersonal situation in which one person must bear a cost for the benefit of the group. If no one volunteers, all suffer. Research shows that many individuals are willing to volunteer, but little is known about the impressions volunteers and defectors make on social perceivers. In three studies, we find that observers judge volunteers to be more competent and more moral than defectors. The outcome of the dilemma, as co-determined by the other person's decision, and the target person's expectations regarding the other's decision systematically affect judgments of competence but have little effect on judgments of morality. Observers' own preferences to volunteer vs. defect influence their judgments of targets consistent with well-known egocentric biases. Taken together, these findings reveal normative social pressures on the potential volunteer, thereby indicating the insufficiency of the conventional game-theoretic analysis.

"We enter the world prepared to perform roles and manage the impressions of others, with the ultimate evolutionary aim of getting along and getting ahead in the social groups that define who we are" (McAdams, 2016).

In social and economic life, people often face a choice between paying a small price in order to benefit the group and doing nothing. If no one steps up to pay that price, all suffer. This type of situation is known as the volunteer's dilemma (VoD; Diekmann, 1985). Many social situations can be modeled as VoDs. In the scholarly literature, the bystander paradigm holds a place of prominence (Fischer et al., 2011). Certain types of emergency require the intervention of one bystander. Each additional volunteer's action is wasted. Just one person needs to offer a seat to the elderly gentleman in the streetcar; only one person needs to call the utility company during a power outage in the neighborhood; only one student needs to answer the professor's question in the lecture hall. Some VoDs are matters of life and death. Who, in times of war, will smother a live grenade to save all others when there is no time to deliberate and coordinate (Blake, 1978)? Whereas the potential losses loom large in these examples, other VoDs highlight the variation in the size of the potential gain. For example, financial windfalls or purchases often require a single 'point person' or coordinator to ensure the transfer of goods. Just one person needs to drive to the pizza shop so that everyone enjoys the meal, and only one designated driver is needed for a small group to enjoy a night out. The framing of the payoffs as gains or losses does not affect the structure of the game, and its impact on decision-making appears to be limited, although it is not nothing (Krueger, Ullrich, & Chen, 2016).

The decision to volunteer (or not) is often made in public. This raises the question: what impressions do volunteers and defectors make on those around them? As McAdams

(2016, see epigraph) observes, much of social judgment and decision-making occurs within a larger context challenging the individual to get along with others while staying competitive at the same time. In the case of volunteering behavior, its moral aspect dominates folk impressions (Schwartz, 1970). Volunteering is readily seen as a prosocial act because it provides a public good. But what of its status as a rational or competent choice? According to some schools of thought (e.g., Plato), the distinction between doing good and acting smart dissolves upon close inspection. Other schools of thought, especially those in the liberal economic tradition, maintain that rationality and competence be defined strictly in terms of self-interest and coherence (Binmore, 2007). According to this view, the morality of choice and behavior is an independent issue at best and a nuisance at worst. In this paper, we are interested in both dimensions of social judgment, morality and competence, which have emerged as a robust framework for social perception (Fiske, Cuddy, & Glick, 2007; Wojciszke, 1994). For each dimension, we examine how individuals are perceived given three types of information or a subset thereof: [i] a target person's choice to volunteer or to defect, [ii] the partner's choice, which co-determines the outcome payoff, and the target person's expectation regarding the other's choice – and thereby the expected outcome. We also ask how our respondents' own preferences to volunteer or defect are related to their judgments of the target persons.

Before reporting the findings of three studies, we review research on the VoD and the social perception of decision-makers in other domains. We then develop and test specific hypotheses regarding social perception in the context of the VoD. We close with a discussion of how decision-making in social dilemmas is interwoven with how such decisions are viewed and evaluated.

**Review of the VoD**

In the VoD, as in other social dilemmas (Dawes, 1980; Krueger, Evans, & Heck, 2017), individuals must choose between behaving prosocially and behaving egocentrically (Diekmann, 1985). Free riding – the egocentric choice – holds the promise of personal rewards as long as others act prosocially. Table 1 shows a gain-framed two-person VoD in matrix form. If both individuals (or 'players,' in game-theoretic parlance) volunteer, each receives a small payoff. If both defect, neither receives anything. If one player defects and the other volunteers, the defector receives a large payoff while the volunteer receives the standard small payoff. In short, volunteering ensures the receipt of a modest gain (or limited loss), whereas unilateral defection makes a larger but uncertain payoff possible while exposing the player to the risk of receiving nothing (or losing a large amount). The dilemma is most acutely experienced in the one-shot, simultaneous, and anonymous version of the game, which precludes coordination through the exchange of promises or other signals. Our research explores this type of situation.

People use a variety of strategies to choose between cooperation and defection (Krueger, Heck, & Wagner, in revision) and rates of volunteering can be high, even to the point of inefficiency (Krueger et al., 2016; Murnighan, Kim, & Metzger, 1993; Przepiorka & Diekmann, 2013). Ideally, a player picks whichever strategy the other is not taking. This goal of complementary coordination distinguishes the VoD from the more familiar prisoner's dilemma (PD). In the PD, defection is the dominating strategy, whereas mutual cooperation is best for the group overall. Here, prosocial players seek positive coordination, that is, they want to make the same prosocial choice. When both cooperate, they achieve Pareto efficiency, that is, the highest possible sum of payoffs. In contrast, the VoD poses the risk of

*over*volunteering because the summed payoffs of two volunteers are lower than what would

be obtained if one person defected (Murnighan et al., 1993; Krueger et al., 2016). The VoD

shares this characteristic with the game of chicken (or hawk-dove game, Rapoport &

Chammah, 1966).

Aside from appeals to prosocial preferences, there are two influential explanations for

why the probability of volunteering can be high. The game-theoretic premise is that players

are rational value maximizers. In a social dilemma without a dominating strategy, such

players volunteer (or cooperate) with a mixed-strategy probability that establishes a Nash

equilibrium if adopted by all (Diekmann, 1985; 1993). For the payoffs displayed in Table 1,

this probability is .5. Another explanation, which is favored by some theorists, works from

the premise that (many) people eschew risk for fear of loss (Novemsky & Kahneman, 2005;

Tversky & Kahneman, 1991). This explanation, inspired by Abraham Wald's (1945)'s early

decision-theoretic work, describes the strategy of choosing that option which minimizes the

largest possible loss. On this view, people approach each available choice option from a

worst-case scenario perspective and choose the one that maximizes the minimum value

(hence the term *maximin* strategy). Some contemporary theories refer to this strategy as the

minimization of anticipated regret (Acevedo & Krueger, 2004; Zeelenberg, 2015). In the

VoD, a volunteer might regret volunteering when learning that the other player also

volunteered, but regret would arguably be greater among mutual defectors who lament the

receipt of the worst of all possible outcomes.

**Social perception**

We submit that, in addition to these game- and decision-theoretic considerations, the

impressions volunteers and defectors make on observers (as well as on each other and

themselves) are an important ingredient of the psychology of choice within the VoD. We further suggest that impressions are also affected by the other player's choice (and hence the outcome for the focal player) and by the first player's expectations regarding the other's choice. In this research, we develop a set of hypotheses and explore them systematically.

Research on social perception has yielded important insights into behavioral decision-making that go beyond a strict game-theoretic analysis (Hartley et al., 2016; Heck & Krueger, 2016; Krueger & Acevedo, 2007). For our studies, we turned to the general model that scales social perception along the two dimensions of competence and morality (Abele, Cuddy, Judd, & Yzerbyt, 2008). In previous work, observer judgments tracked target behavior in social dilemmas and their self-presentations. For example, defectors in the PD are perceived as more competent and less moral than cooperators (Krueger & Acevedo, 2007, Krueger & DiDonato, 2010). This work also demonstrated an outcome bias in perceptions of competence. An outcome bias exists when a person is not only judged in light of the soundness of the decision process, but also on the basis of specific outcomes that were unknowable at the time of decision (Baron & Hershey, 1988; Cushman, Dreber, Wang, & Costa, 2009; Zelazo, Helwig, & Lau, 1996). To illustrate, the quality of a bet should be judged only with respect to its expected value and the expected values of bets not taken. The bet should not be considered good or bad if it was won or lost. Showing outcome bias in the PD, observers judge unilateral cooperators as less competent than mutual cooperators, although the players did not know the other's choice at them time of their own decision – hence the dilemma (Krueger & Acevedo, 2007). Applied to the VoD, the social-perception approach allows us to test whether perceivers are influenced by the dilemma's outcome,

which is co-determined by the target person's partner, and whether they are influenced by the target person's expectations regarding the partner's decision.

Expectation biases have been demonstrated such that observers are sensitive to the coherence (or lack thereof) of target persons' decisions and expectations. It would seem incompetent, for example, to choose an improbable payoff over a probable payoff of the same size. In the PD, observers dislike conditional defectors, that is, players who defect against those whom they expect to cooperate (Krueger & DiDonato, 2011). Such individuals seem exploitative and hence immoral.

We now lay out a set of hypotheses concerning the effects of target's behavior in the VoD, their outcomes, and their expectations of others' choices. Over the three studies, our predictions differ for the domains of competence and morality; thus, our experimental hypotheses are detailed separately for each.

**Hypotheses**

**Perceived competence**

We advance four hypotheses for perceptions of competence. Our first hypothesis refers to the effect of the target person's decision. We predict that volunteers are seen as more competent than defectors. The rationale for this hypothesis is that the VoD allows individuals to avoid risk by volunteering. They can make ensure that the worst outcome (Player 1 defects and Player 2 defects; D:D) will not be obtained. Inasmuch as respondents are themselves loss or risk-averse, they may credit volunteers with positive judgments of competence. This prediction is cast against the background of the alternative possibility is that volunteers and defectors are seen as equally competent, because neither has violated a dominating strategy. With a Nash equilibrium of .5, volunteering with the same probability

of a flipped coin turning up 'heads' satisfies game-theoretic rationality and may therefore render volunteering and defecting in a single dilemma equally competent.

Our second hypothesis refers to the effect of the other person's decision on judgments of the target person. The partner's choice co-determines the target person's outcome and may thus be seen as relevant. If there is outcome bias, an individual who defects will be seen as lacking in competence if their partner also defects and thus receives the lowest payoff. This prediction follows from the finding that outcome bias takes a negative form in the prisoner's dilemma (Krueger & Acevedo, 2007). In the PD, individuals who are defected against are seen as less competent than individuals whose outcome is unknown. It is also critical to note that in the VoD (in contrast to the PD), a volunteer's own outcome does not vary as a result of the partner's choice (although the partner's outcome does). Put together, the first two hypotheses amount to the following pattern of judged competence in descending order: V:V ~ D:V > V:D > D:D. In the language of analysis of variance (ANOVA), this pattern comprises a main effect of target decision, a main effect of partner decision, and an interaction between the two.

Our third hypothesis addresses the interplay of a person's decision and that person's expectation regarding the partner's choice. We predict that competence ratings will be lower if targets do not follow the best course of action suggested by their own expectations. Those who act against their own perceived interest will be seen as less competent than those who do not. In the VoD, a rational individual chooses the strategy opposite the one expected from the partner. Competent individuals volunteer if they expect their partners to defect, and defect if they expect their partners to volunteer. The outcomes associated with these choices may also matter. Specifically, when the target's own decision is the opposite of the expected partner

decision, the defecting target person has the best outcome. Among the two cases where own decision and expected partner decision are the same, the defecting target person reaps the lowest outcome. These considerations result in the following pattern of predicted competence ratings: $D:V_{expected} > V:D_{expected} \sim V:V_{expected} > D:D_{expected}$. Analytically, this pattern amounts to a main effect of expected partners' choice and an interaction between expectation and the target's own choice.

Our fourth hypothesis considers the question of accuracy, that is, the effects of the partner's expected choice and the partner's actual choice, and thus the outcome of the dilemma. We predict that those targets whose partners choose as expected will be seen as more competent than targets whose partners violate expectations. Perceived competence should be higher for those targets who make accurate predictions about their partners' decisions. When expectations are violated, the cost to the target varies. A target who expects volunteering but gets defection experiences a costly betrayal, and may be seen as less competent than a target who expects defection but gets volunteering (a windfall). This effect would amount to a more complex kind of outcome bias. The ordering of predicted competence ratings is therefore: $V_{expected}:V_{received} \sim D_{expected}:D_{received} > D_{expected}:V_{received} > V_{expected}:D_{received}$, and should manifest itself in a main effect of other's choice and an interaction effect.

In theory, there might be grounds for greater complexity, such that there might be a three-way interaction between the target's choice, the other's choice, and the target's expectation regarding the other's choice. Lacking a sufficient theoretical foundation, we refrained from predicting a specific pattern for such an interaction, and left this analysis exploratory.

**Perceived morality**

Thanks to past research and *a priori* considerations of plausibility, our hypotheses regarding perceptions of morality are straightforward. We predict that morality judgments depend primarily on the target's decision to volunteer or defect. This hypothesis derives from research showing the dominance of the morality dimension in person perception (Goodwin, Piazza, & Rozin, 2014; Strohminger & Nichols, 2014) and research in the prisoner's dilemma (Krueger & Acevedo, 2007) and social comparison (Heck & Krueger, 2016). In both domains, perceptions of morality show no evidence of outcome or expectation bias.

**Egocentrism**

Finally, we explore the contribution of respondents' own choice preference to their perceptions of others. There are three specific predictions. First, we assume that respondents' will claim a greater willingness to volunteer compared with the average person. This prediction derives from the idea that volunteering is a form of valued (moral) prosocial behavior and the general finding that self-enhancement biases are particularly strong in the moral domain (Alicke & Govorun, 2005; Tappin & McKay, 2016). Second, we predict that respondents' own willingness to volunteer predicts their estimates of the likely behavior of the average person. This prediction derives from extensive research on social projection (see Robbins & Krueger, 2005, for a meta-analysis). Third, and most intriguingly, we explore the novel idea that observers will judge a player as competent and moral inasmuch as that person responds to the VoD as they themselves would. In other words, we submit that social judgment in the domain of the VoD has an egocentric element (see Epley, Converse, Delbosc, Monteleone, & Cacioppo, 2009, for related findings).

**A Note on Methods and Analysis**

We conducted three studies concurrently but present them in order of increasing complexity for ease of exposition. A prospective power analysis (Faul, Erdfelder, Lang, & Buchner, 2007) showed that for a sample of 90 participants per condition ($\alpha = .05$), we could detect an effect of $d = .26$ in a repeated measures comparison, and an effect of $d = .37$ in an independent groups comparison, with a probability of .80. Effects of this magnitude (or larger) have been observed in past work on the VoD (Krueger et al, in revision) and social perception (Heck & Krueger, 2016; Krueger & Acevedo, 2007; Krueger & DiDonato, 2011). We collected data from 100 participants in studies 1 and 2, and 200 participants for Study 3, which had two conditions. We estimated that 10% of participants might have to be excluded for failing attention probes. Each respondent was allowed to participate in no more than one of these studies.

We examined reliability of observer judgments for the entire sample. With 1,230 observations[1], we found that the short scales for competence (intelligent, rational, naïve (reverse scored)) and morality (ethical, trustworthy, selfish (reverse scored)) had satisfactory reliability (mean inter-item correlations = .52 [$\alpha = .75$] and .59 [$\alpha = .79$], respectively). These results did not differ substantially between studies, and are consistent with previous research (e.g., Krueger & Acevedo, 2007). The two scales were moderately correlated over respondents within and between conditions, $r(1,228) = .47$. To rule out shared variance or suppressor effects, we also performed all analyses with repeated covariates (competence controlling for morality and vice versa) (ANCOVA, Tabachnik & Fidell, 2007, pp. 214-215). Doing so changed only one statistical inference compared with the simpler ANOVA

---

[1] Study 1, $N = 89$, two target observations per participant; Study 2, $N = 88$, four target observations per participant; Study 3, $N = 175$, four target observations per participant.

approach; we discuss it in Study 3[2]. We therefore proceed with the standard ANOVA results.

In each study, we excluded participants who checked the scale midpoint for every rating.

This exclusion rule was based on the assumption that these participants had disengaged from

the task. No participants provided uniform responses for ratings other than the scale

midpoint. The data of 9 participants were dropped (2.6% of the total sample). Including these

data in the analyses did not change the direction or significance of any statistical effect.

### Study 1: Perceptions of Behavior

The goal of the first study was to see if judgments of competence and morality are

sensitive to the target person's choice between volunteering and defection. This study yields

a baseline measure of how people perceive each decision (volunteer, defect) in the absence of

information about outcomes or expectations.

### Method

Participants ($N = 100$) were recruited on Amazon Mechanical Turk (MTurk; Amazon,

2014) after being screened using TurkGate (2013) to ensure that they had not previously

participated in our studies on the VoD. Eligibility for participation was restricted to residents

of the United States. Participants received $0.45 as compensation. Average completion time

was 5:56 minutes.

After providing consent, participants were told that they would be asked to rate a

series of individuals who had made a decision in a social dilemma. The next page offered a

description of the VoD, with a presentation of all four possible outcomes and a payoff matrix

visualizing the structure of options and outcomes. The outcome structure (see Table 1) was

adapted for display in U.S. dollar amounts ($0.00, $25.00, or $50.00). Participants were told

---

[2] Raw data, SPSS syntax, and the data structure used for ANCOVA analyses are available as Supplemental
Materials.

that the two players could not communicate with each other and that their decisions would be revealed simultaneously. Targets were labeled "Player 1" and "Player 2," and their choices were labeled "Option A" and "Option B." The terms "volunteer" and "defect" were not used at any point. Participants were given three chances to correctly answer two comprehension questions about the structure of the dilemma before proceeding.

Participants then provided ratings of two separate, androgynously named targets: one who had chosen "Option A," and one another who had chosen "Option B." These options corresponded to volunteering and defecting, respectively. One example target description follows: *"Jesse recently played as Player 1 in the game we have described to you. Jesse chose Option A, and does not know what option Player 2 chose. The possible outcomes are marked in red."* A payoff matrix accompanied this description with a red box highlighting the choice and outcomes for "Option A."

Each target was presented on a separate page, the order of which was counterbalanced. Ratings were made for six trait adjectives comprising scales of competence (Intelligent, Rational, Naïve (reverse scored)) and morality (Ethical, Trustworthy, Selfish (reverse scored)) on scales ranging from 1 (Not at all) to 5 (Extremely). Trait adjectives were presented in alphabetical order within each scale and presented in the same order for each target. Following this primary social perception task, participants also rated a brief series of four unrelated targets for a separate study. At the end of the experiment, participants were asked to provide their own preference for Option A in the VoD game, and their estimate of the average Mechanical Turk user's preference for Option A on a five point scale. Finally, participants provided demographic information and were debriefed.

**Results and Discussion**

Ten participants failed the comprehension check and another one checked the scale midpoint for every rating. The data of these individuals were excluded from analysis, leaving a sample of 89.

Target ratings were entered into a 2 (behavior: defect/volunteer) by 2 (domain: competence/morality) repeated measures ANOVA. Means are displayed in Figure 1. As predicted by our first hypothesis, a volunteer was perceived overall more positively than a defector, $F_{Behavior}(1, 88) = 76.02$, $p < .001$, $\eta_p^2 = .46$, but there was also an main effect of domain, $F_{Domain}(1, 88) = 7.36$, $p = .008$, $\eta_p^2 = .08$, and a significant behavior-by-domain interaction, $F_{Behavior*Domain}(1, 88) = 5.32$, $p < .001$, $\eta_p^2 = .46$. This interaction reflects the finding that the target's choice had a greater effect on morality judgments than on competence judgments. Simple effects showed that volunteers were seen as more competent than defectors, $F(1, 88) = 50.65$, $p < .001$, $\eta_p^2 = .37$, $d = .76$[3], 95% CI of the difference [.61, 1.01], and more moral, $F(1,88) = 66.87$, $p < .001$, $\eta_p^2 = .43$, $d = .87$, 95% CI of the difference [.86, 1.41].

Because volunteering was seen more positively than defection in both domains, and knowing that people tend to overestimate their own desirable and ethical behaviors (Alicke, 1985; Brown, 2011; Epley & Dunning, 2000; Tappin & McKay, 2016) we predicted that participants would claim to be more likely to volunteer than others. Indeed, there was a clear better-than-average effect. Participants reported that they themselves were more likely to choose Option A (volunteer) ($M = 4.65$, SD = 1.50) than the average Mechanical Turk user ($M = 3.93$, SD = 1.43), $t(88) = 3.95$, $p < .001$, $d = .42$, 95% CI of the difference [.36, 1.08].

---

[3] This and all subsequent repeated measures Cohen's $d$ values were calculated using the formula for correlated measures, $d = ((M1 - M2) / SD) / \sqrt{(1 - r)}$ (Cohen, 1988).

At the same time, the correlation between these two types of rating over participants was also significant, $r_{Self,Other}(87) = .32$, $p = .001$, which corroborates the projection hypothesis. In short, perceivers believed that others would choose as they themselves would, while being less inclined to volunteer overall. Although this finding is also consistent with self-stereotyping, such an interpretation is not likely in a domain where no clear stereotypes exist (Van Veelen, Otten, Cadinu, & Hansen, 2015). Even when self-referent judgments follow other-referent judgments in time (as in this study), a correlation between the two most likely signals social projection. In order to make other-referent judgments, respondents tend to bring self-related information or preferences to mind first, a process that may occur automatically (Dunning & Hayes, 1996; Krueger, 2003).

Next, we tested the novel egocentrism hypothesis to see if respondents' own preferences to volunteer predicted their perceptions of volunteers and defectors. We centered self- and other-ratings and entered them along with their cross-products into a multiple regression model to predict the difference between ratings of volunteering and ratings of defection as the criterion. Indeed, the more observers were themselves inclined to volunteer, the more competent, $\beta = .419$, $t(85) = 4.17$, $p < .001$, and the more moral, $\beta = .424$, $t(85) = 4.06$, $p < .001$, they rated volunteers as compared with defectors. Conversely, and consistent with the egocentrism hypothesis, respondents' predictions of the average Mechanical Turk user's preference to volunteer had no such effects, $\beta_{Competence} = .058$, $t(85) = .576$, $p = .566$, $\beta_{Morality} = -.085$, $t(85) = -.821$, $p = .414$. The interaction term similarly failed to contribute to the model, $\beta_{Competence} = -.145$, $t(85) = -1.50$, $p = .138$, $\beta_{Morality} = -.079$, $t(85) = -.782$, $p = .436$.

To review, the results of Study 1 supported the hypotheses regarding the target person's choice of strategy in the VoD and regarding the influence of the observers' own

preferences. Study 2 was designed to test a broader set of hypotheses. Beyond providing a concurrent replication of the preliminary findings, Study 2 allowed a test of the second hypothesis, that is, the idea that outcome bias would selectively affect perceptions of competence.

### Study 2: Perceptions of Behavior and Outcome

Participants ($N = 100$) were recruited and screened. The procedures for Study 2 were similar to the ones used in Study 1 with one exception: the target descriptions now also included the choices made by Player 2, thereby revealing the monetary outcome of the dilemma for both players. Participants rated each of four targets (Player 1) who found themselves in a situation of mutual volunteering, mutual defection, or unilateral defection favoring either Player 1 or Player 2. In each scenario, the outcome was marked with a red box indicating how much money each target would receive as a result of both players' decisions (see Appendix for an example stimulus). Descriptive text was presented below this image summarizing the choices each player made. An example of this descriptive text for the mutual volunteering case reads: *"Taylor recently played as Player 1 in the game we have described to you. Taylor chose Option A. It turns out the person Taylor played the game with chose Option A. This outcome is marked in red."* The order of the four targets was randomized for each participant.

### Results and Discussion

Nine participants failed the comprehension check and three other participants selected the scale midpoint for every rating. The data of these individuals were excluded from analysis, yielding a total sample size of 88.

**Judgments of behavior**

We first revisited the pattern observed in Study 1 (Hypothesis 1) by collapsing over the choices made by Player 2 and entering ratings of volunteering and defecting targets into a 2 (behavior: defect/volunteer) by 2 (domain: competence/morality) repeated measures ANOVA. As in Study 1, respondents rated volunteers more favorably than defectors, $F(1, 87) = 120.14$, $p < .001$, $\eta_p^2 = .58$. The effect was of medium size for competence, $F(1, 87) = 63.36$, $p < .001$, $\eta_p^2 = .42$, $d = .85$, 95% CI of the difference [.588, .980], and strong for morality, $F(1, 87) = 109.18$, $p < .001$, $\eta_p^2 = .56$, $d = 1.12$, 95% CI of the difference [1.61, 1.10]. Both the main effect of domain, $F(1,87) = 5.42$, $p = .022$, $\eta_p^2 = .06$, and the behavior-by-domain interaction, $F(1, 87) = 21.86$, $p < .001$, $\eta_p^2 = .20$, were replicated.

Consistent with the prediction of self-enhancement bias, participants again claimed that they were more likely to volunteer ($M = 4.69$, SD = 1.57) than the average respondent ($M = 3.84$, SD = 1.49), $t(87) = 4.99$, $p < .001$, $d = .53$, 95% CI of the difference [.51, 1.19]. Consistent with the idea of social projection, these two types of rating were positively correlated, $r(86) = .46$, $p < .001$. In support of the egocentrism hypothesis, participants' own preference to volunteer predicted the difference between competence ratings of volunteering and defecting behavior, $\beta = .353$, $t(84) = 3.03$, $p = .003$, and morality ratings, $\beta = .444$, $t(84) = 3.82$, $p < .001$. The more observers were themselves inclined to volunteer, the more favorably they rated volunteering targets as compared with defecting targets. Other-estimates had no such effect, $\beta_{Competence} = .060$, $t(84) = .531$, $p = .597$, $\beta_{Morality} = -.053$, $t(84) = -.478$, $p = .634$, nor did the interaction terms, $\beta_{Competence} = -.057$, $t(84) = -.540$, $p = .591$, $\beta_{Morality} = .023$, $t(84) = .217$, $p = .829$.

**Judgments of behavior and outcome**

We next tested for outcome bias (Hypothesis 2) with separate 2 (behavior: volunteer/defect) by 2 (outcome: Player 2 volunteers/defects) repeated-measures ANOVAs for competence and morality ratings (see Figure 2). Volunteers were rated as more competent than defectors, $F(1, 87) = 63.36$, $p < .001$, $\eta_p^2 = .42$. This was not surprising, given the replication of the pattern observed in Study 1. The finding of interest was the main effect of other's choice, $F(1, 87) = 61.63$, $p < .001$, $\eta_p^2 = .415$, such that targets were perceived as less competent when Player 2 defected rather than volunteered. This bias was stronger when targets themselves defected than when they volunteered, as indicated by the significant interaction term, $F(1, 87) = 12.29$, $p < .001$, $\eta_p^2 = .124$. Simple effects analyses showed that participants judged volunteers as less competent when the other player defected rather than volunteered, $F(1, 87) = 20.05$, $p < .001$, $\eta_p^2 = .187$, $d = .48$, 95% CI of the difference [.46, .18], and also that they judged defectors as less competent when their partner defected, $F(1, 87) = 55.06$, $p < .001$, $\eta_p^2 = .388$, $d = .79$, 95% CI of the difference [.50, .87]. Perceptions of competence thus tracked target behavior, but were also sensitive to the dilemma's outcome. An act of defection by Player 2 resulted in lower competence ratings of the target.

Judgments of morality showed a strikingly different pattern (see Figure 2, bottom panel). Here, we found the familiar effect of behavior such that volunteering was perceived as more moral than defection, $F(1, 87) = 109.18$, $p < .001$, $\eta_p^2 = .557$, but there was no outcome bias. Neither the main effect, $F(1, 87) = 2.45$, $p = .121$, $\eta_p^2 = .027$, nor the behavior-by-outcome interaction were statistically significant, $F(1, 87) = .13$, $p = .722$, $\eta_p^2 = .001$. In the moral domain, perceptions were sensitive only to targets' decision to volunteer or defect.

**Judgments of behavior and outcome relative to baseline (Study 1) judgments of**

**behavior**

When competence judgments are sensitive to outcome information, it is instructive to

compare outcome-based perceptions with baseline ratings made in the absence of outcome

information. Accordingly, we conducted planned comparisons between the samples in Study

1 and Study 2 (see Figure 2, dashed lines for baseline means obtained from Study 1).[4]

Relative to a volunteering baseline target ($M = 3.83$, SD $= .74$; Study 1), there was no

difference in perceived competence when compared with a target who volunteered together

with Player 2 (Study 2), ($M = 3.85$, SD $= .61$), $t(175) = .20$, $p = .420$, $d = .03$, 95% CI of the

difference [-.18, .22]. However, a volunteer who met with defection ($M = 3.53$, SD $= .71$),

was perceived as less competent than this baseline, $t(175) = 2.71$, $p = .007$, $d = .41$, 95% CI

of the difference [.08, .51]. In other words, the outcome bias affecting perceptions of

volunteers was selectively negative. Being the object of defection entails a loss of ascribed

competence.

We conducted similar comparisons for defectors. Relative to the perceived

competence of a baseline defector (Study 1) ($M = 2.98$, SD $= .76$), defectors whose partner

volunteered (Study 2) were seen as more competent, $t(175) = 3.69$, $p < .001$, $d = .55$, 95% CI

of the difference [.19, .64]. Achieving unilateral defection was viewed as more competent

than simply choosing to defect. Those targets who were defected against, resulting in

undesirable mutual defection, were perceived as less competent ($M = 2.57$, SD $= .73$) than

baseline defectors, $t(175) = 2.45$, $p = .015$, $d = .37$, 95% CI of the difference [.05, .49]. For

---

[4] Cross-study comparisons must be viewed with caution. In the present case, we trust these comparisons because the studies were run concurrently.

the defectors, then, outcome bias was bi-directional, bringing a benefit or a loss depending on the other person's choice.

Judgments of morality yielded no differences between baseline and fully described targets, all $p$'s > .10.

### Study 3: Perceptions of Behavior, Outcome, and Expectation

Study 3 provided another opportunity to assess the consistency of the first set of findings. In addition, this study introduces tests of the hypotheses referring to the target person's expectations (Hypotheses 3 and 4). Participants ($N = 200$) were recruited and screened as in the first two studies. Targets' expectations were added as a between-subjects variable to the design. Participants were randomly assigned to a condition where each of four targets expected their partner to volunteer or to a condition where each target expected the other person to defect. To provide information regarding targets' expectations, we added a sentence to the description noting that "[target] expected Player 2 to choose Option A (Option B).' The image displaying the four possible outcomes and their payoffs to each player contained an additional line of text clarifying that "Player 1 expects Player 2 to choose Option A (Option B)." An example target description of mutual volunteering in the 'volunteering expected' condition follows: *"Taylor recently played as Player 1 in the game we have described to you. Taylor expected that Player 2 would choose Option A. Taylor chose Option A. It turns out the person Taylor played the game with chose Option A. This outcome is marked in red."* All other materials, including the dependent measures, were identical to those in Study 2.

**Results and Discussion**

Twenty-one participants failed the comprehension check and four additional

participants answered every question by choosing the scale midpoint. After excluding the

data of these participants, 84 participants remained for analysis in the 'volunteering

expected' condition and 91 remained in the 'defection expected' condition.

Before describing the statistical effects involving the target's expectations of

volunteering vs. defection, we pooled the data over the two levels of the expectation variable

and replicated the analyses performed in Study 2. Figure 3 shows that the patterns were much

the same; replication analyses can be found in the supplemental materials. With the basic

patterns secured, we proceeded to ask whether expectations moderated these results.

Adding the between-subjects variable of expectation to the ANOVA model resulted

in a 2 (behavior: volunteer/defect) by 2 (outcome: partner volunteers/defects) by 2

(expectation: target expects partner to volunteer/defect) repeated measures design with one

between-subjects factor (expectation), separately for competence and morality (see Figure 3).

The three-way interactions were not significant for competence, $F(1, 173) = 2.48$, $p = .120$,

or morality $F(1, 173) = .93$, $p = .340$, suggesting that the critical pattern of results did not

depend on targets' expectations of their partner.

**Competence**

Targets who expected volunteering were perceived as more competent than targets

who expected defection, $F(1, 173) = 89.83$, $p < .001$, $\eta_p^2 = .342$. We had not foreseen this

result. Perhaps respondents viewed volunteering as normative in both the descriptive and the

injunctive sense (Murnighan, et al. 1993), and therefore judged those targets who expected

normative behavior as more competent. These targets expressed expectations consistent with the social norm.

To test our third hypothesis, we examined the behavior-by-expectation interaction. The interaction term was significant, $F(1, 173) = 18.75$, $p < .001$, $\eta_p^2 = .098$, but its direction differed from the predicted one (see Figure 3, top panels). Perceived competence depended more on that target's own behavior when defection rather than volunteering was expected. Simple effects analyses showed that volunteers were perceived as similarly competent whether they expected their partner to volunteer ($M = 3.52$, SD = .59) or defect ($M = 3.63$, SD = .58), $F(1, 173) = 1.37$, $p = .246$, $d = .18$, 95% CI of the difference [-.28, -.07]. These targets avoided the costly outcome of mutual defection even if they were defected against. As predicted, however, defectors were perceived as less competent if they also expected their partner to defect, ($M = 2.66$, SD = .64), rather than volunteer ($M = 3.16$, SD = .64), $F(1, 173)$ $= 26.63$, $p < .001$, $d = .79$, 95% CI of the difference [.31, .69] (see Figure 3, top panels, rightmost bars). We think this finding may stem from the recognition that mutual defection is the worst outcome for all. Interestingly, targets who expected their partner to volunteer were perceived as more competent if they also volunteered ($M = 3.52$, SD = .52) rather than defected ($M = 3.16$, SD = .60), $F(1, 83) = 16.52$, $p < .001$, $d = .45$, 95% CI of the difference [.18, .54] (Figure 3; top left panel). It appears that *not* choosing to defect against someone one expects to volunteer was viewed as more rational despite the relative inefficiency of mutual volunteering.

Testing our fourth hypothesis, our primary interest was in the outcome-by-expectation interaction. As expected, respondents judged those targets favorably who correctly predicted their partner's choice, $F(1, 173) = 7.32$, $p = .0071$, $\eta_p^2 = .04$. The pattern

of this interaction was partially consistent with the hypothesis. Targets who expected

volunteering were perceived as more competent when their expectation turned out to be

correct ($M = 3.64$, SD = .47) rather than incorrect ($M = 3.05$, SD = .55), $F(1, 83) = 61.03$, $p <$

.001, $d = .88$, 95% CI of the difference [.44, .73] (Figure 3; top left panel). Targets who

expected defection, however, were perceived as more competent when their expectation was

incorrect ($M = 3.29$, SD = .54) rather than correct ($M = 3.00$, SD = .58), $F(1, 83) = 12.50$, $p <$

.001, $d = .38$, 95% CI of the difference [.13, .45] (Figure 3; top right panel). Here, the effect

of the other player's decision (i.e., the game's outcome) overrode the effect of a target

making an expectation error. Of the two possible errors a target can make in the VoD, a false

expectation of volunteering is more serious because it can result in the worst outcome of

mutual defection. Being able to anticipate defection is the most important task in the VoD.

Yet, targets who correctly expect defection were viewed as less competent than targets whose

partners volunteer regardless of their own expectation.

**Morality**

Whereas the actual and the expected choice of player 2 moderated perceptions of

competence, they had little effect on perceptions of morality (Figure 3; bottom panels). There

was no main effect of expectation, $F(1, 173) = .001$, $p = .98$, $\eta_p^2 = .000$. Of the interaction

effects, only the outcome-by-expectation effect was significant, $F(1, 173) = 4.62$, $p = .033$,

and it was small, $\eta_p^2 = .026$. Those who expected volunteering were viewed as less moral

when their partner defected against them ($M = 3.12$, SD = .37) than when their partner

volunteered ($M = 3.28$, SD = .43), $t(83) = 3.44$, $p = .010$, $d = .38$, 95% CI of the difference

[.07, .25]. Conversely, those who expected defection were seen as similarly moral regardless

of whether their partner volunteered ($M = 3.21$, SD = .45) or defected ($M = 3.19$, SD = .48),

$t(90) = .56, p = .580, d = .38$, 95% CI of the difference [-.06, .11]. However, this effect was

no longer significant when controlling for perceived competence.

The two remaining interaction effects were not significant, $F_{outcome*behavior}(1, 173) =$

$1.03, p = .311$; $F_{behavior*expectation}(1,173) = 2.03, p = .156$. When including perceived

competence as a covariate, the behavior-by-expectation interaction achieved significance,

suggesting a greater degree of difference in perceived morality between targets who expected

volunteering and targets who expected defection. Here, expecting volunteering and choosing

to defect was seen as the least moral behavior.[5]

To summarize, the target's behavior, expectation, and the partner's actual decision

(outcome), jointly shape perceptions of that target's competence. In contrast, perceptions of

morality depend almost exclusively on the target's choice between volunteering and

defecting.

## General Discussion

In three studies, we find consistent patterns of how people perceive and judge

individuals in a VoD. The general finding is that they judge volunteers to be both more

competent and more moral than defectors. This broad effect may help explain why many

people volunteer even when the rewards are modest (or costs are great). Knowing that others

view volunteers as positively as they themselves do, people may use opportunities for

providing a public good in a strategic effort to get along and get ahead at the same time

(McAdams, 2016; see epigraph).

Yet, social perceptions depend on more than what a person (the target) does. The

second finding is that outcome bias erodes perceptions of competence. Outcome bias violates

---

[5] We also re-tested the self-enhancement and social-projection hypotheses. The findings replicated those
obtained in Studies 1 and 2. The details are in the Supplemental Materials.

the rational norm that only expected (*ex ante*) consequences affect a decision, but not

revealed (*ex post*) consequences. Observers should not judge a person by the consequences of

their decisions, but only by how the person valued and weighted these consequences at the

time of decision-making. Consistent with earlier research on this bias (Baron & Hershey,

1988; Krueger & Acevedo, 2007), victims of defection were perceived as comparatively less

competent. Outcome bias did not affect perceptions of morality.

The third finding is that respondents impose a penalty on those individuals who

defect while having correctly predicted that their partner would also defect against them.

That is, the reputational damage is greatest for those individuals who knowingly act against

their own (and the other's) best interest. One possible interpretation of this finding, which

remains to be tested, is that observers respond to what they see as a spiteful termination of an

ultimatum game (Güth, 1995). The target person who believes the partner will defect has a

choice between accepting the modest payoff for volunteering while leaving the prize to the

defector, and rejecting the deal with the result that neither person receives anything.

Rejecting the ultimatum is also the incompetent response according to game theory.

Finally, consistent with an extensive literature in social cognition, perceptions in the

VoD are biased by observers' own preferences. Respondents show the typical biases of self-

enhancement and social projection, and they judge target persons positively inasmuch as

these persons responded to the VoD as they themselves would. We now review the patterns

of judgment separately for the domains of competence and morality.

**Competence**

Why do observers think it is more competent to volunteer than to defect? Observers

may recognize that volunteering is the only way to eliminate the most aversive outcome of

mutual defection. Observers might also feel that individuals who volunteer are sending out a

costly but effective signal that they care about public goods. The construction of a reputation

as a prosocial person may signal social competence (Barclay, 2004; Griskevicius, Tybur, &

Van den Bergh, 2010). If volunteering predicts a person's likelihood to volunteer in a similar

situation, it is reasonable for volunteering behavior to be perceived as the competent choice

even in an anonymous, one-shot environment. Because successfully navigating the VoD

requires interpersonal coordination, merely observing that a target is willing to volunteer may

be enough for observers to consider her a competent player.

We predicted that outcome bias would moderate perceptions of competence. Being

defected against may be seen as an indication of lacking competence, even if by the lights of

strict rationality outcome information should be ignored (Baron & Hershey, 1988). Observers

may have believed that targets who were defected against ultimately deserved it (Lerner,

2003). Some observers may have thought that a competent person would anticipate the other

person's choice. Such counterfactual explanations are not sufficient to absolve observers

from the charge of bias. Participants knew that the targets were engaged in a one-shot,

anonymous dilemma and that the two players could not communicate with each other. Still,

those who were defected against were seen as suckers. Would these biased perceptions

persist if Player 2 had instead been a random number generator? This possibility remains to

be tested.

Study 2 produced an unanticipated result: targets who achieved mutual volunteering

(V:V) were perceived as more competent than targets who successfully defected in the

presence of a volunteer (D:V) (see Figure 2). In the latter case, the target obtained a large

payoff without harming Player 2. However, observers felt that the volunteer was more

competent despite the inefficient outcome, that is, the missed opportunity to take advantage

of the other's volunteering. This result replicated in Study 3 under stricter conditions. Here,

volunteers were seen as more competent than defectors who expected their partner to

volunteer. This result runs counter to the hypothesis that the shrewdest individuals (defectors

expecting volunteering) would be perceived as more competent than inefficient but prosocial

others (volunteers expecting volunteering). We speculate that observers infer high

competence in prosocial targets because they expect these targets to find themselves in future

situations where prosociality will lead to favorable outcomes – a halo effect. Alternatively, it

is conceivable that respondents used a simple 'prosociality is good' heuristic, causing them to

underweight the targets' expectations of others and instead focus on the behavior itself.

**Morality**

Whereas perceptions of competence were biased by outcomes and targets'

expectations, perceptions of morality tracked only a target's decision to volunteer or defect.

There was no evidence for outcome or expectation effects; volunteering (defection) was

consistently rated above (below) the scale midpoint. Recent evidence suggests that

perceptions of morality are less about wealth and fitness and more about character,

trustworthiness, and the 'essence' of self (Goodwin et al., 2014; Hartley et al., 2016;

Strohminger & Nichols, 2014). In light of these findings, it makes sense that only a target's

behavior would determine perceptions of their morality.

We attribute the perceived immorality of defection to the implied willingness to

expose others to risk for the sake of own potential gain. Defection in the VoD amounts to a

claim of power because it ensures that one's own outcome cannot be worse than the other's.

Consider a situation in which Player 2 knows Player 1's decision. Such a sequential VoD

would amount to an ultimatum game. Player 2 is left with only a single viable option (volunteer), where the negative alternative is to ensure no gain for either player by defecting. Withholding agency from Player 2 is likely to elicit disapproval.

Negative information looms large in moral judgments (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Rozin & Royzman, 2001), whereas positive information is most critical in the competence domain (Reeder & Brewer, 1979; Skowronski & Carlston, 1989). This asymmetry suggests that one observed instance of defection relative to several cases of volunteering may be enough to justify giving lower morality ratings (Epley & Klein, 2016). To the extent that moral judgments are about perceptions of essence, it is easy to infer that a defector is seen as morally corrupt or tainted (Rozin & Nemeroff, 1990), thus reflecting their likely immorality in the future (i.e., 'spoiling the pot').

To be clear, we do not suggest that morality judgments never vary with outcomes or expectations. Research on victim-blaming and rape culture (Niemi & Young, 2014), as well as punishment in cases of accidental or misperceived harm (Cushman et al., 2009; Gray, Waytz, & Young, 2012), militate against this suggestion.

**Conclusion**

We have presented evidence for the idea that social perceivers view the decision to volunteer as both competent and moral, while being biased by a dilemma's resolution only when judging competence. We suggest that rational interests, including loss-aversion, reputation management, and coordination signaling, combine to explain why so many people choose to volunteer. Because defection is viewed unfavorably relative to volunteering overvolunteering can occur, particularly among individuals who are socially close (Krueger et al., 2016). Overvolunteering (i.e., mutual volunteering) is a Type I error or wasted

investment, whereas undervolunteering (mutual defection) is a Type II error or missed

opportunity to do good (Heck & Krueger, 2015; Swets, Dawes, & Monahan, 2000). The

former error is less grievous than the latter, which may explain the overall bias – among

players and their observers – toward volunteering.

The volunteer's dilemma poses unique challenges to social decision makers and the

scientists who study them. If the 'game' were played optimally and consistently (e.g., by

flipping a coin), the ability to coordinate decisions in uncertain environments may have failed

to develop. A social cognitive system designed to identify and reward prosocial individuals

through praise and reputation must exist if we are to learn how to coordinate successfully

with one another.

**References**

Abele, A. E., Cuddy, A. J., Judd, C. M., & Yzerbyt, V. Y. (2008). Fundamental dimensions of social judgment. Special Issue of the *European Journal of Social Psychology*. doi: 10.1002/ejsp.574

Acevedo, M., & Krueger, J. I. (2004). Two egocentric sources of the decision to vote: The voter's illusion and the belief in personal relevance. *Political Psychology, 25,* 115-134. doi:10.1111/j.1467-9221.2004.00359.x

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology*, *49*, 1621–1630. doi:10.1037/0022-3514.49.6.1621

Alicke, M. D., & Govorun, O. (2005). The better-than-average effect. In M. D. Alicke, D. Dunning & J. Krueger (Eds.), The self in social judgment (pp. 85-106). New York: Psychology Press.

Barclay, P. (2004). Trustworthiness and competitive altruism can also solve the "tragedy of the commons". *Evolution and Human Behavior*, *25*, 209-220. doi: 10.1016/j.evolhumbehav.2004.04.002

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, *54*, 569-579. doi: http://dx.doi.org/10.1037/0022-3514.54.4.569

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*, 323-370. doi: http://dx.doi.org/10.1037/1089-2680.5.4.323

Binmore, K. (2007). *Game theory: A very short introduction*. Oxford, England: Oxford University Press.

Blake, J. A. (1978). Death by hand grenade: Altruistic suicide in combat. *Suicide and Life-Threatening Behavior*, *8*, 46-59. doi: 10.1111/j.1943-278X.1978.tb01084.x

Brown, J. D. (2012). Understanding the better than average effect: Motives (still) matter. *Personality and Social Psychology Bulletin*, *38*, 209–219. doi:10.1177/0146167211432763

Cohen, J. (1988). Statistical power analysis for the social sciences. Hillsdale, NJ, Erlbaum.

Cushman, F., Dreber, A., Wang, Y., & Costa, J. (2009). Accidental outcomes guide punishment in a "trembling hand" game. *PloS one*, e6699. doi: http://dx.doi.org/10.1371/journal.pone.0006699

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology, 31*, 169-193. doi: 10.1146/annurev.ps.31.020180.001125

Diekmann, A. (1985). Volunteer's dilemma. *Journal of Conflict Resolution, 29*, 605-610. doi: 10.1177/0022002785029004003

Diekmann, A. (1993). Cooperation in an asymmetric volunteer's dilemma: Theory and experimental evidence. *International Journal of Game Theory, 22*, 75-85. doi: 10.1007/BF01245571

Dunning, D., & Hayes, A. F. (1996). Evidence for egocentric comparison in social judgment. Journal of Personality and Social Psychology, *71,* 213-229. doi: http://dx.doi.org/10.1037/0022-3514.71.2.213

Epley, N., Converse, B. A., Delbosc, A., Monteleone, G. A., & Cacioppo, J. T. (2009). Believers' estimates of God's beliefs are more egocentric than estimates of other people's beliefs. *Proceedings of the National Academy of Sciences*, *106*, 21533-21538. doi: 10.1073/pnas.0908374106

Epley, N., & Dunning, D. (2000). Feeling "holier than thou": Are self-serving assessments

produced by errors in self-or social prediction? *Journal of Personality and Social*

*Psychology*, *79*, 861-875. doi:10.1037/0022-3514.79.6.861

Evans, A. M., & Krueger, J. I. (2011). Elements of trust: Risk taking and expectation of

reciprocity. *Journal of Experimental Social Psychology*, *47*, 171-177.

doi:10.1016/j.jesp.2010.08.007

Evans, A. M., & Krueger, J. I. (2014). Outcomes and expectations in dilemmas of trust.

*Judgment & Decision Making, 9*, 90-103. doi:

http://journal.sjdm.org/13/13502/jdm13502.html

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* power 3: A flexible

statistical power analysis program for the social, behavioral, and biomedical

sciences. *Behavior Research Methods*, *39*, 175–191. doi: 10.3758/BF03193146

Fischer, P., Krueger, J. I., Greitemeyer, T., Vogrinic, C., Kastenmüler, A., Frey, D., Heene,

M., Wicher, M., & Kainbacher, M. (2011). The bystander effect: A meta-analytic

review on bystander intervention in dangerous and non-dangerous emergencies.

*Psychological Bulletin, 137*, 517-537. doi: 10.1037/a0023304

Fiske, S. T., & Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition:

warmth and competence. *Trends in Cognitive Sciences, 11*, 77-83.

doi: http://dx.doi.org/10.1016/j.tics.2006.11.005

Goldin, G., Darlow, A. (2013). TurkGate (Version 0.4.0) [Software]. Available from

http://gideongoldin.github.com/TurkGate/

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person

perception and evaluation. *Journal of Personality and Social Psychology, 106*, 148-

168. doi: 10.1037/a0034726

Gray, K., Waytz, A., & Young, L. (2012). The moral dyad: A fundamental template unifying

    moral judgment. *Psychological Inquiry*, *23*, 206-215. doi:

    10.1080/1047840X.2012.686247

Griskevicius, V., Tybur, J. M., & Van den Bergh, B. (2010). Going green to be seen: status,

    reputation, and conspicuous conservation. *Journal of Personality and Social*

    *Psychology*, *98*, 392-404. doi: http://dx.doi.org/10.1037/a0017346

Güth, W. (1995). On ultimatum bargaining experiments – a personal review. *Journal of*

    *Economic Behavior & Organization, 27,* 329-344. http://dx.doi.org/10.1016/0167-

    2681(94)00071-L

Hartley, A. G., Furr, R. M., Helzer, E. G., Jayawickreme, E., Velasquez, K. R., & Fleeson, W.

    (2016). Morality's centrality to liking, respecting, and understanding others. *Social*

    *Psychological and Personality Science*, online first. doi: 10.1177/1948550616655359.

Heck, P. R., & Krueger, J. I. (2015). Self-enhancement diminished. *Journal of Experimental*

    *Psychology: General, 144,* 1003-1020. doi: http://dx.doi.org/10.1037/xge0000105

Heck, P. R., & Krueger, J. I. (2016). Social perception of self-enhancement bias and error.

    *Social Psychology, 47,* 327-339. doi: 10.1027/1864-9335/a000287

Klein, N., & Epley, N. (2016). Maybe holier, but definitely less evil, than you: Bounded self-

    righteousness in social judgment. *Journal of Personality and Social Psychology*, *110*,

    660-674. doi: http://dx.doi.org/10.1037/pspa0000050

Krueger, J. I. (2003). Return of the ego—self-referent information as a filter for social

    prediction: Comment on Karniol (2003). *Psychological Review, 110*, 585-590.

    doi:10.1037/0033-295X.110.3.585

Krueger, J. I., & Acevedo, M. (2007). Perceptions of self and other in the prisoner's dilemma:

Outcome bias and evidential reasoning. *American Journal of Psychology, 120*, 593-

618.

Krueger, J. I., & DiDonato, T. E. (2010). Perceptions of morality and competence in (non)

interdependent games. *Acta Psychologica, 134*, 85-93.

doi:10.1016/j.actpsy.2009.12.010

Krueger, J. I., Evans, A. M., & Heck, P. R. (2017). Let me help you help me: Trust between

profit and prosociality. In P. A. M. Van Lange, B. Rockenbach, & T. Yamagishi

(Eds.). *Social dilemmas: New perspectives on trust.* New York: Oxford University

Press.

Krueger, Heck, & Wagner. (in revision). *Egocentrism in the volunteer's dilemma*. Brown

University.

Krueger, J. I., Ullrich, J., & Chen, L. J. (2016). Expectations and decisions in the volunteer's

dilemma: effects of social distance and social projection. *Frontiers in Psychology:

Cognition, 7,* article 1909. doi: 10.3389/fpsyg.2016.01909

Lerner, M. J. (2003). The justice motive: Where social psychologists found it, how they lost it,

and why they may not find it again. *Personality and Social Psychology Review, 7*,

388-399. doi: 10.1207/S15327957PSPR0704_10

McAdams, D. P. (2016, June). The mind of Donald Trump. *The Atlantic*.

http://www.theatlantic.com/magazine/archive/2016/06/the-mind-of-donald-

trump/480771/

Murnighan, J. K., Kim, J. W., & Metzger, A. R. (1993). The volunteer dilemma.

*Administrative Science Quarterly, 38*, 515-538. doi: 10.2307/2393335

Niemi, L., & Young, L. (2014). Blaming the Victim in the Case of Rape. *Psychological Inquiry*, *25*, 230-233. doi: 10.1080/1047840X.2014.901127

Novemsky, N., & Kahneman, D. (2005). The boundaries of loss aversion. *Journal of Marketing Research, 42*, 119-128. doi: http://dx.doi.org/10.1509/jmkr.42.2.119.62292

Przepiorka, W., & Diekmann, A. (2013). Individual heterogeneity and costly punishment: a volunteer's dilemma. *Proceedings of the Royal Society of London B: Biological Sciences*, *280*. 20130247. doi: 10.1098/rspb.2013.0247

Rapoport, A., & Chammah, A. M. (1966). The game of chicken. *American Behavioral Scientist, 10*, 10-28. doi: 10.1177/000276426601000303

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review*, *86*, 61-79. doi: http://dx.doi.org/10.1037/0033-295X.86.1.61

Robbins, J. M., & Krueger, J. I. (2005). Social projection to ingroups and outgroups: A review and meta-analysis. *Personality and Social Psychology Review*, *9*, 32-47. doi: 10.1207/s15327957pspr0901_3

Rozin, P., & Nemeroff, C. (1990). The laws of sympathetic magic: A psychological analysis of similarity and contagion. In J.W. Stigler, R.A. Shweder, & G. Herdt, (Eds.*), Cultural psychology: Essays on comparative human development*. (pp. 205-232). New York, NY: Cambridge University Press.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*, 296-320. doi: 10.1207/S15327957PSPR0504_2

Schwartz, S. H. (1970). Elicitation of moral obligation and self-sacrificing behavior: an experimental study of volunteering to be a bone marrow donor. *Journal of Personality and Social Psychology*, *15*, 283-293. doi: http://dx.doi.org/10.1037/h0029614

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, *105*, 131-142. doi: http://dx.doi.org/10.1037/0033-2909.105.1.131

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition, 131*, 159–171. doi: 10.1016/j.cognition.2013.12.005

Swets, J., Dawes, R., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*, 1–26. doi: 10.1111/1529-1006.001

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson Education.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The quarterly journal of economics*, 1039-1061.

Van Veelen, R., Otten, S., Cadinu, M., & Hansen, N. (2015). In integrative model of social identification: Self-stereotyping and self-anchoring as two cognitive pathways. *Personality and Social Psychology Review, 20*, 3-26. doi: 10.1177/1088868315576642

Wald, A. (1945). Statistical decision functions which minimize the maximum risk. *Annals of Mathematics,* 46, 265-280. doi: 10.2307/1969022

Wojciszke, B. (1994). Multiple meanings of behavior: Construing actions in terms of competence of morality. *Journal of Personality and Social Psychology, 67*, 222-232. doi: http://dx.doi.org/10.1037/0022-3514.67.2.222

Zeelenberg, M. (2015). Robust satisficing via regret minimization. *Journal of Marketing Behavior, 1*, 157-166. doi: 10.1561/107.00000010

Zelazo, P. D., Helwig, C. C., & Lau, A. (1996). Intention, act, and outcome in behavioral prediction and moral judgment. *Child Development*, *67*, 2478-2492. doi: 10.2307/1131635

Table 1

*Structure of the Volunteer's Dilemma*

Column Player

| | | Volunteer | Defect |
|---|---|---|---|
| Row | Volunteer | 1 \| 1 | 1 \| 2 |
| Player | Defect | 2 \| 1 | 0 \| 0 |

*Figure 1*. Study 1. Scale means for competence and morality ratings of volunteering and defecting targets. Error bars represent one standard error of the mean.
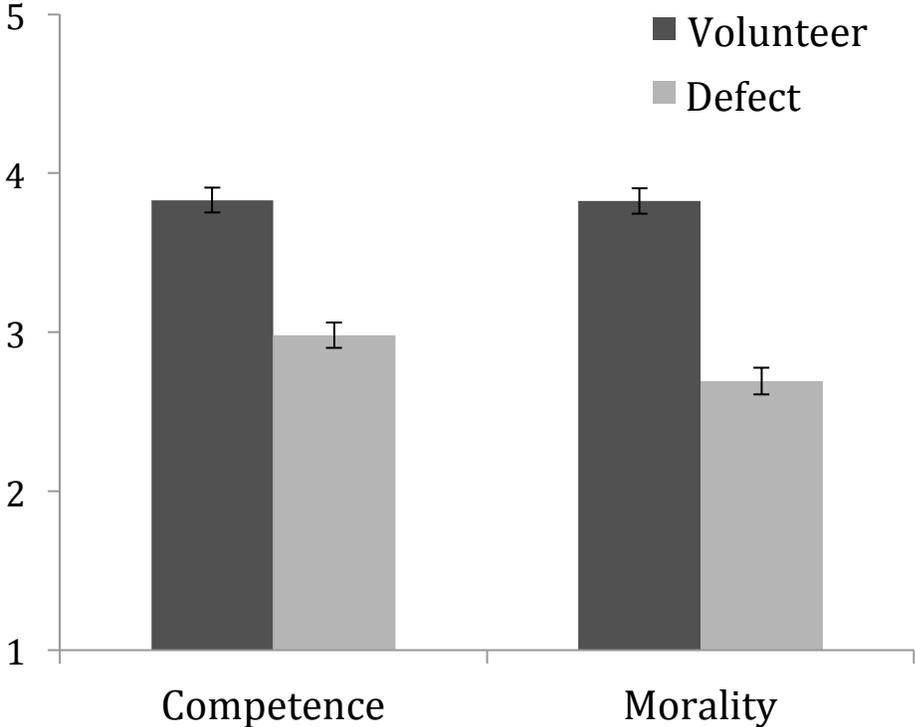
*Figure 2.* Study 2. Scale means for competence and morality ratings of volunteering and defecting targets whose partners chose to volunteer or defect. Dashed columns display baseline means observed in Study 1 for volunteering and defecting targets.  Error bars represent one standard error of the mean.
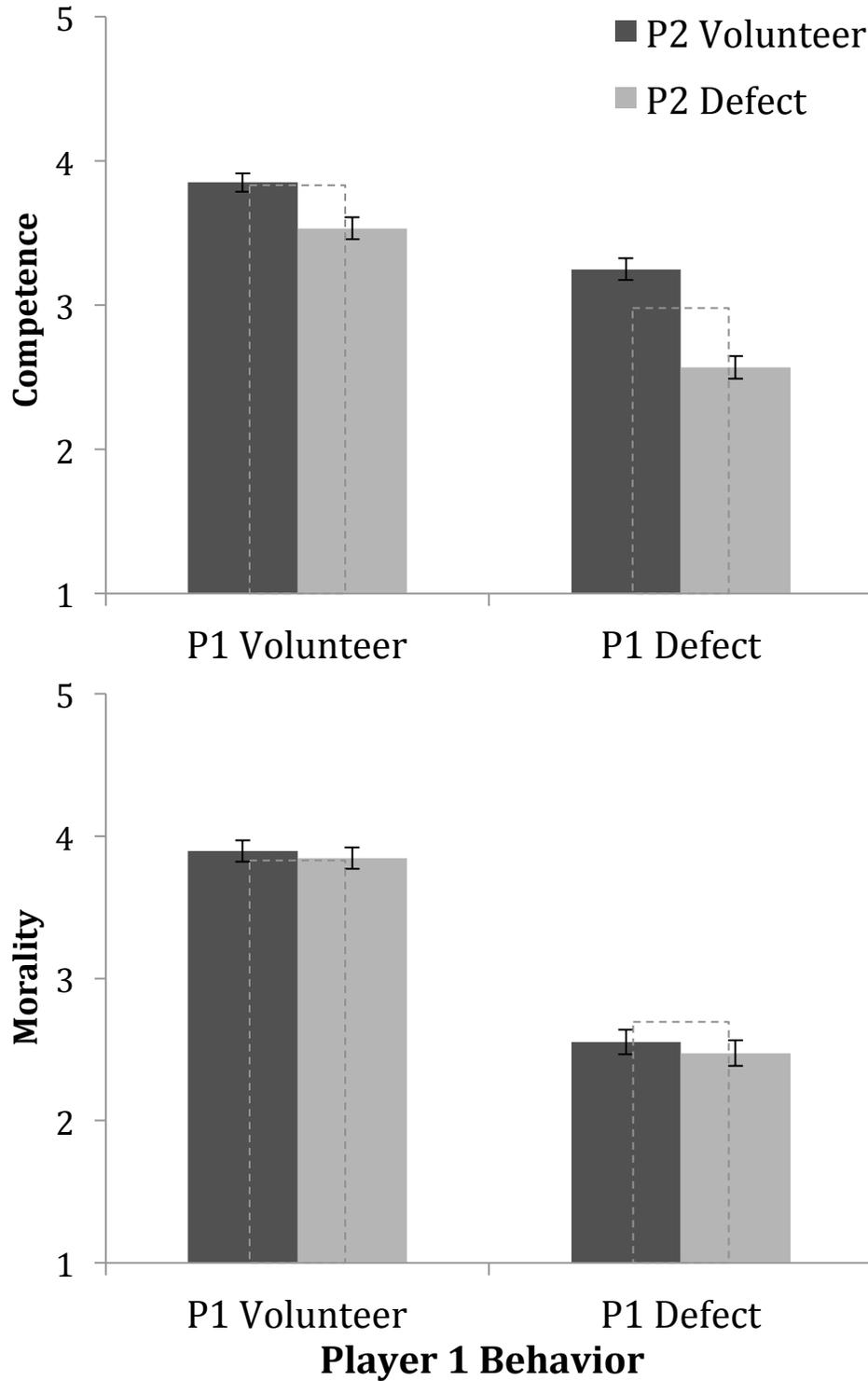
*Figure 3.* Study 3. Scale means for competence and morality ratings of volunteering and defecting targets expecting either volunteering or defection. Dashed columns display baseline means observed in Study 1 for volunteering and defecting targets.  Error bars represent one standard error of the mean.
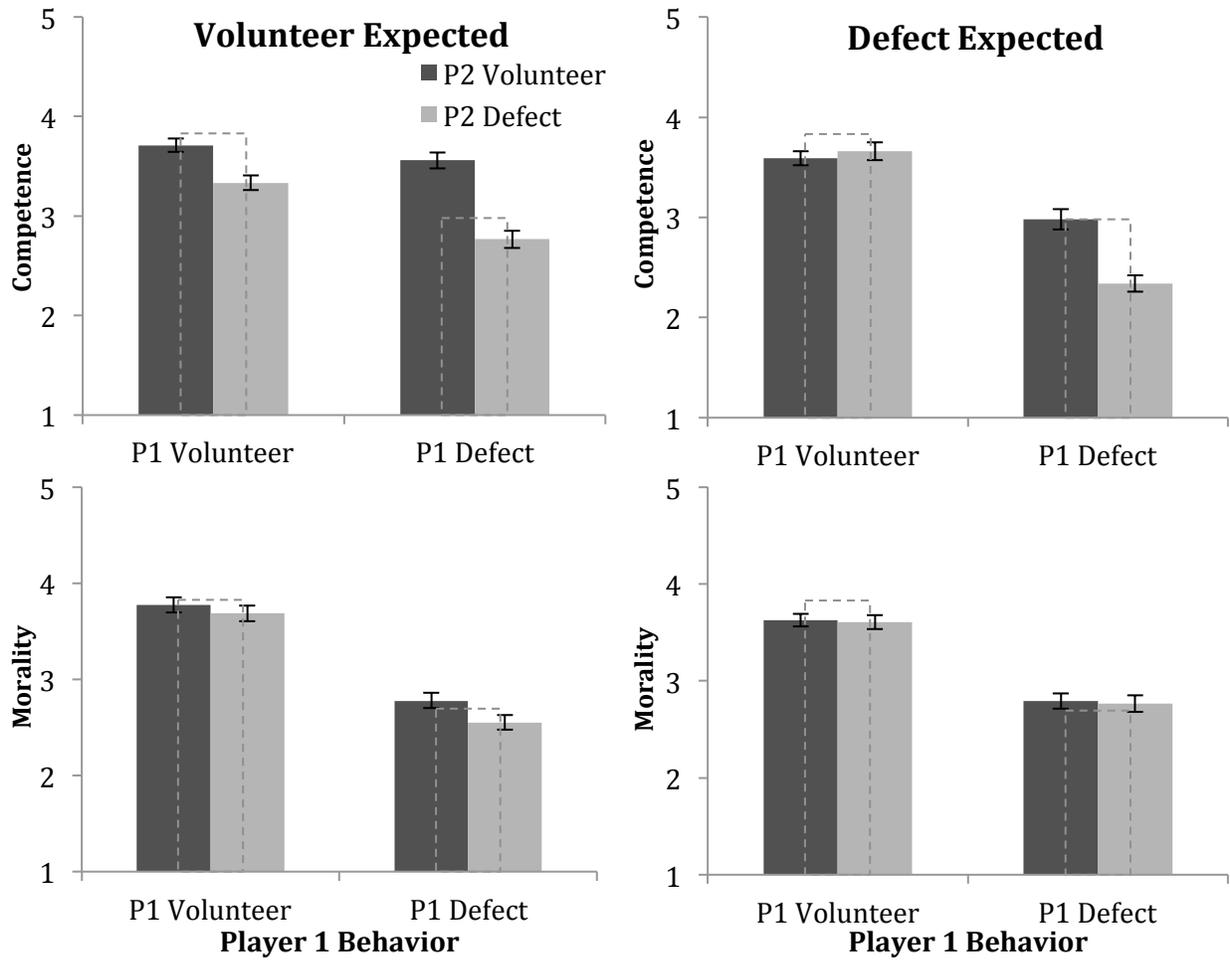
Appendix

*Figure A1.* Screenshot of the accompanying image presented to participants alongside target description text. This example was taken from Study 2 for a target who chose Option A and whose partner chose Option A. Note that the rectangle indicating the outcome for both players was displayed to participants in red.