Altruism and Indirect Reciprocity:

The Interaction of Person and Situation in Prosocial Behavior

Brent Simpson
University of South Carolina

and

Robb Willer
University of California, Berkeley

December 2007

Total Word Count = 8,600

* This research was supported by grants SES-0551895 and SES-0647169 from the *National Science Foundation*. We thank Nick Berigan, Rachael Chatterson, Lisa Dilks, Alex Hirschfeld, Jonathan Hyndman, Kyle Irwin, Tucker Mcgrimmon, and Jenny Triplett for valuable research assistance. We also thank Pat Barclay, Kimmo Eriksson, Detlef Fetchenhauer, and Matt Salganik for helpful comments on previous drafts. Direct correspondence to Brent Simpson, Department of Sociology, University of South Carolina, Columbia SC 29208 or e-mail bts@sc.edu

Altruism and Indirect Reciprocity:

The Interaction of Person and Situation in Prosocial Behavior

Abstract

A persistent puzzle in the social and biological sciences is the existence of *prosocial behavior*, actions that benefit others, often at a cost to oneself. Recent theoretical models and empirical studies of indirect reciprocity show that actors behave prosocially in order to develop an altruistic reputation and receive future benefits from third parties. Accordingly, individuals should stop investing in reputations via prosocial behavior when a future benefit (via indirect reciprocity) is unlikely. The conclusion that the absence of reputational incentives necessarily leads to egoistic behavior contrasts sharply with models of heterogeneous social preferences. Such models demonstrate the theoretical plausibility of populations composed of egoists *and* altruists. Results of Study 1 show that actors classified *a priori* as egoists respond strategically to reputational incentives, whereas those classified *a priori* as altruists are less affected by these incentives. While egoists act prosocially when reputational incentives are at stake but not when opportunities for indirect reciprocity are absent, altruists tend to act prosocially regardless of whether reputational incentives are present. These results suggest that apparently altruistic behavior can result from non-strategic altruism or reputation-building egoism. Study 2 demonstrates the robustness of these results and explores indirect reciprocation of others' prosocial acts. The results show that altruists indirectly reciprocate at higher levels than egoists, and individuals tend to discount others' prosocial behaviors when they occur in the presence of reputational incentives. As a result, public prosocial behaviors are indirectly reciprocated less than private prosocial behaviors. In line with our argument that altruists pay less attention to reputational incentives, egoists showed a greater tendency than altruists to discount others' public prosocial behaviors. The results support the growing focus on heterogeneity of individuals' social preferences in models of altruism and indirect reciprocity.

Altruism and Indirect Reciprocity:
The Interaction of Person and Situation in Prosocial Behavior

Humans display a wide array of *prosocial behaviors*, actions that benefit others, often at a cost to oneself. For decades, scholars have drawn on theories of kin selection (Hamilton 1964) and reciprocal altruism (Trivers 1971) to explain such actions. But much prosocial behavior is directed at unrelated strangers who may be unable to reciprocate at some later time, which presents an anomaly for both explanations. To explain prosocial behavior among unrelated individuals in large populations, researchers turned their attention to models of *indirect reciprocity* (Alexander 1987; Nowak 2006; Panchanathan and Boyd 2004; Rockenbach and Milinski 2006; Wedekind and Milinski 2000), which account for the possibility that individuals receive long-term benefits for short-term prosocial acts.

Both field work (Smith and Bird 2000) and laboratory studies (Milinski et al. 2002a; 2002b) support the notion that individuals confer important advantages on those who act prosocially toward others, i.e., they reciprocate benefactors indirectly. Furthermore, research shows that potential benefactors respond strategically to the presence or absence of these benefits, cooperating at higher levels when reputational benefits and possibilities of indirect reciprocity exist versus when they do not (Milinski et al. 2002b; Semmann et al. 2004).

The presence or absence of opportunities for indirect reciprocity thus may explain much of prosocial behavior. But here we build on existing models of heterogeneous social preferences to show that indirect reciprocity approaches miss a very important piece of the prosociality puzzle and, as a result, under-predict the occurrence of prosocial behavior when opportunities for indirect reciprocity do not exist. Specifically, we show that whereas some individuals (here referred to as “egoists”) behave prosocially only when reputational incentives encourage the

behavior, others ("altruists") behave prosocially without significant regard for reputational incentives. Thus, taking account of heterogeneity of social preferences – the idea that some actors hold largely egoistic preferences while others possess more altruistic preferences – promises a much clearer understanding of the dynamics of prosocial behavior.

Heterogeneity of actors' other-regarding preferences has played a critical role in many theoretical models of prosocial behavior (Bowles and Gintis 2004; Dugatkin and Wilson 1991) but empirical studies are needed (Kurzban and Houser 2005). As others (Fehr and Fischbacher 2005) have argued, empirical studies of indirect reciprocity that do address heterogeneity of social preferences generally do so *ex-post*, using the dependent measure of interest to classify individuals as egoists or altruists. Because it can lead to unfalsifiable claims, such approaches greatly cripple the power of explanations that rely on preference heterogeneity. The method we use here classifies individuals a priori as *egoists* or *altruists* and thus avoids the unfalsifiability of a post hoc procedure for classification.

The arguments tested in Study 1 of this paper rest at the intersection of research on indirect reciprocity and heterogeneous social preferences. As such, they explain prosociality based on a *person x situation* interaction. Specifically, egoists are expected to act prosocially only when they anticipate receiving rewards or avoiding costs for doing so. Altruists, on the other, behave prosocially because they are motivated to increase others' welfare, even at a cost to their own payoffs. Thus, as explained more fully below, we expect that egoists will be affected by reputational incentives, i.e., whether future benefactors will observe or will learn about the prosocial act. As a result, egoists will behave prosocially when reputational incentives are at stake but not when opportunities to benefit from "indirect reciprocity" are absent. Altruists, on

the other hand, will behave prosocially regardless of whether reputational incentives exist or not. We test these predictions in our first study.

In Study 2 we turn our attention to the indirect reciprocation of others' prosocial behaviors. This part of our research addresses three related questions. First, do indirect reciprocators "discount" prosocial behavior that occurs in the presence of ulterior motives? That is, do individuals indirectly reciprocate *less* when reputational incentives might have encouraged giving? Second, do altruists reward others' giving more than do egoists? Third, do altruists and egoists exhibit different tendencies to discount prosocial behavior in their patterns of indirect reciprocation? We suggest that altruists and egoists differ in the extent to which they attribute others' giving to ulterior motives (e.g., the pursuit of reputational benefits). If so, this should show up in how individuals reward or indirectly reciprocate others' "private" and "public" prosocial behaviors.

The remainder of the paper is organized as follows. After clarifying key terms, we briefly review research on indirect reciprocity. Thereafter, we link indirect reciprocity models with work on heterogeneous social preferences to explain who acts prosocially when reputational incentives exist versus when they do not. We then present the results of two laboratory experiments designed to test key predictions against alternative predictions suggested by previous work. Finally, we conclude with suggestions for future work.

Terminology

We define prosocial behavior as any behavior that benefits another person, often at a cost to the benefactor. Altruism is the motivation to increase another's welfare and egoism is the motivation to increase one's own welfare. We use the labels "altruism" and "egoism" largely out of necessity. There is very little agreement within most literatures on the definition of altruism

(or related terms like prosocial behavior), and even less agreement across literatures, including the two lines bridged by our work (see Batson 1987 for a discussion). We settled on these labels because, as shown in detail below, they fit with common-sense notions of altruism and egoism. Altruists (in our arguments and in our experimental results) act prosocially without anticipating future direct (or indirect) reciprocation, whereas egoists act prosocially when they anticipate reciprocation (in this case, indirect reciprocation). We will therefore distinguish between prosocial behaviors motivated by altruism (altruistic behavior) and egoism (egoistic behavior).

As explained in detail below, our research measures egoistic versus altruistic motivations and values using standard measures. Unlike previous research in the indirect reciprocity literature, which observes prosocial behavior in the absence of any obvious egoistic gains and then explains that behavior *ex post* as altruistic, we follow the social value orientation literature: we measure values *a priori* and use this measure to predict altruistic versus egoistic prosocial behaviors.

Indirect Reciprocity

As noted earlier, the classic explanations of prosocial behavior are kin selection (Hamilton 1964) and reciprocal altruism (Trivers 1971). But these theories do not explain prosocial behaviors directed at non-kin who may be unable or unwilling to reciprocate. For instance, what explains the benevolence of a person who donates to a charity for needy children or others who cannot readily offer a *quid pro quo*? Recent theories and research on *indirect reciprocity* (Alexander 1987) address exactly these types of questions. This research advances the notion that prosocial behaviors are rewarded with, and motivated by, reputational benefits. Whereas direct reciprocity approaches describe the tendency of individuals to directly return favors to

others who have helped them in the past, indirect reciprocity occurs when an individual is rewarded for being generous to someone else in the past.

A number of theories account for indirect reciprocity processes (for a recent review, see Nowak 2006). For instance, theories of *costly signaling* (Zahavi 1995) assume that actors use conspicuous behaviors to signal fitness or quality. Bystanders are then expected to find those who use costly signals as more attractive mates or exchange partners. Similarly, theories of *image scoring* (Nowak and Sigmund 1998) and *image standing* (Sugden 1986; Leimar and Hammerstein 2001; Panchanathan and Boyd 2004) point to how prosocial acts lead to reputational benefits. A central tenet of these models is that those with more prosocial reputations accrue benefits through indirect reciprocity.

As noted earlier, empirical work has supported two basic predictions from theories of indirect reciprocity: First, individuals confer important advantages on those who act prosocially toward others, i.e., they reciprocate indirectly. Second, potential benefactors respond strategically to the presence or absence of these reputational incentives. That is, they cooperate at higher levels when these incentives are present than when they are not.

Indirect Reciprocity and Heterogeneous Social Preferences

The indirect reciprocity framework is useful for understanding prosocial behavior. But, much like kin selection and reciprocal altruism approaches, we believe the framework suffers from a key limitation: a homogenous view of the actor. To our knowledge, indirect reciprocity approaches invariably assume a consistently self-interested actor. As Semmann et al. (2004, 248-49) put it, these theories predict that "individuals should stop investing in costly reputation as soon as they find out that a future pay-off is unlikely to occur."

In this section, we extend the indirect reciprocity approach with a model of the actor that allows for heterogeneous social preferences. Our predictions assume two theoretically pure types of actors, one egoistic (heretofore assumed in indirect reciprocity approaches) and one altruistic. Egoists consider only their own payoffs (either via direct or indirect reciprocity) when deciding whether to act prosocially, doing so only if prosociality produces benefits for themselves. Altruists, on the other hand, consider the payoffs to the beneficiary of a potential prosocial act.[1]

Hypotheses

Person x *Situation Hypothesis*: We predict an interaction between the person (egoist/altruist) and situation (the presence/absence of reputational incentives). Specifically we expect altruists will give more than egoists when there are no reputational benefits for doing so, *i.e.,* in private. Egoists, on the other hand, will seek to appear altruistic when there are reputational benefits (in public). Thus, we expect little or no difference between the behaviors of altruists and egoists in the public compared to the private condition.

However, this is not the only possible pattern of results. Previous work suggests a number of competing predictions.

Alternative 1: Main Effect of Situation: Previous research on indirect reciprocity predicts a main effect of situation such that all individuals give more in public settings, where reputational incentives are present. Because this line of research assumes a homogeneous view of the actor, from this perspective, we would not expect to see an effect of the person.

[1] A number of assumptions are possible for the relative weight altruists assign to their own and others' outcomes (see Van Lange 1999). We begin by assuming that altruists assign equal weight to own and a potential beneficiary's payoff. We leave exploration of alternative assumptions for future research.

Alternative 2: Main Effect of Person: Similarly, because previous research on heterogeneous social preferences does not account for the role of indirect reciprocity, this perspective leads us to predict a main effect of the person, but no effect for the situation.

Alternative 3: Insincere Altruism: Alternatively, we could observe a person *x* situation interaction effect different from the one outlined in our main hypothesis. Specifically, it may be that altruists, rather than egoists, respond positively to reputational incentives. This would be expected from the dominant theoretical perspectives in the biological and economic sciences, which traditionally have held more cynical views on altruistic motivation, such as the sociobiologist who famously quipped "scratch an altruist, watch a hypocrite bleed" (Ghiselin 1974). From this perspective, we would expect (so-called) altruists to give only when reputational incentives encourage them to (i.e. not in private). Meanwhile, egoists would behave selfishly in public and in private. The experiment to follow was designed to test our person *x* situation hypothesis, as well as these alternative specifications.

Study 1: Prosocial Behaviors Depend on a Person *x* Situation Interaction

Design

Participants were recruited from introductory classrooms at a large public university in the southeastern U. S. for the opportunity to earn money. A total of 89 students (50.6% female), ages 18-46 participated. We used a series of "dictator games" to present participants with opportunities to act prosocially. There was a between-subjects factor (egoist or altruist) and a within-subjects factor (participant's decision in the dictator game was public or private). To assess the first factor, we administered a standard "social value orientation" inventory designed to classify each participant as an egoist or altruist (Liebrand et al. 1986; Van Lange 1999). For the second factor, we presented each participant with two opportunities to help a dependent

other. In one case (the private condition), there were no reputational incentives for helping. In a second case (the public condition) there were reputational incentives to give. We randomly counter-balanced the ordering of the private and public conditions.

Procedure

Participants were scheduled in groups of six to eight. Upon entering the laboratory, each participant was escorted to a private subject room. After reading and completing a consent form, participants were presented with instructions that stated they would not see other participants at any point during or after the study, and that participants would be identified only via letters.

Social Value Orientations. Participants completed the "triple dominance" measure of social value orientation (Van Lange 1999). Social value orientations are stable preferences for how outcomes are distributed between self and others. For instance, "egoists" seek to maximize their own outcomes without regard for the outcomes of other players. This is the standard assumption made in most applications of rational choice and game-theoretic models of decision-making and, often, the social sciences in general. "Altruists" give equal weight to their own and others' outcomes (see, e.g., Van Lange 1999). (The social value orientation literature generally uses the label *prosocials*, rather than altruists. Given the awkwardness of distinguishing between "egoists" and "prosocials" who act "prosocially," here we use the term *altruists*.)

The measure of social value orientation presents participants with a series of nine decomposed games, each consisting of three different distributions of points for self and another (unidentified) person. The resulting classifications are *egoism* (which maximizes the payoff to self, without regard to the payoff of the other), *competitivism* (maximizes the difference between the payoff to self and other) and *altruism* (maximizes the aggregate payoff to self and other). Previous research has established the ecological validity (Bem and Lord 1979), temporal stability

(Van Lange 1999), and overall predictive power (Liebrand 1986) of such measures for a wide range of behaviors.

Given the small number of "competitors" in the general population, researchers typically combine these with egoists (e.g., De Cremer and Van Lange 2001). We follow that approach here. Following previous work (Van Lange 1999) we classify participants as a given social value orientation only if they make at least six out of nine choices consistent with a given social value orientation (*egoism* or *altruism*). Based on these criteria, we were able to classify eighty out of 89 participants as having a particular social value orientation (51 altruists, 29 egoists).

Prosocial Behavior Opportunities: After completing the social value orientation measures, participants were presented with a series of opportunities to act prosocially. Our measure of prosociality is behavior in a "dictator game." The dictator game is composed of two players, the "Dictator" (referred to in the study as the "Decider") and "Receiver." The Dictator is given a pool of resources (e.g., $8). His or her task is to decide how much, if any, of this pool of resources to pass on to the Receiver. Any amount not sent is kept by the Dictator. For instance, for a pool of eight resource points each worth $1, if a Dictator decides to send three, the Receiver earns a total of $3 and the Dictator receives $5. Thus, consistent with our definition of prosocial behavior, the Dictator faces an entirely voluntary decision to allocate valued resources to the Receiver, or keep them for him/herself. As shown in Table 1, the dictator games were broken down into two phases (public and private), each ostensibly containing two decision scenarios. In the first decision scenario of each phase, the participant was the Dictator.

[Table 1 about here]

In the second scenario of each phase, a new (third) participant ostensibly decided how much of a $12 resource pool to send to the participant. To keep the behavior of others from influencing

participants' decisions (and to guarantee all participants a reasonable payoff), we simulated the choices of others. In all cases, participants were told that the Dictator in the second decision scenario of each phase decided to send ½ of the resource pool ($6).

The two phases differed according to whether the participant's dictator decision (in the first decision scenario) was "private" or "public." In the private condition, only the participant (Dictator) and Receiver recipient would know the participant's decision. But for the first decision scenario in the public condition, the participant's decision was fictively relayed to a third participant who would ostensibly be the dictator (and the participant the recipient) in the second decision scenario of that particular phase.

In both scenarios (private and public) for which the participant was the dictator, the participant was given $8 to divide. Also in both scenarios the pool size in the second phase was $12. This created an incentive for egoists to develop a reputation for being prosocial in the public behavior condition, as greater giving in phase 1 could elicit greater receiving in phase 2 (Milinski et al. 2002a; 2002b). Because participants were told that decisions were not relayed to third parties, no such reputational benefits existed in the private condition. Each participant made a decision in the public and private condition. The order of conditions was randomly determined for each participant.

After both decisions were made, participants were paid (between $18 and $28) and thoroughly debriefed. In the debriefing session, we probed for suspicion and none was reported. The entire procedure took about 35 minutes.

Results

Descriptive Statistics. Experiment 2 (reported below) uses a different resource pool size. Thus, to increase comparability across experiments, the Figures display the amounts returned as

proportion of the resource pool size. Allocations by egoists and altruists were similar in the public condition, where the modal response for both was to give half. But very different behavioral tendencies emerged in the private condition, where there were no reputational benefits to be gained. As in the public condition, the modal response of altruists was to give half. But, for egoists, the modal response was to keep the entire pool of resources for themselves. Figure 1 displays the average proportion returned by situation (public versus private) and person (egoism versus altruism).

[Figure 1 about here]

To test our hypotheses, we conducted a repeated measures ANOVA, with our predictor variables (situation and person), and the order of conditions (whether participants made a decision in the public or private condition first). The overall *F* tests show a significant effect of person [$F(1, 76) = 11.33$, $p = .001$], situation [$F(1, 76) = 56.69$, $p \leq .001$], and the person *x* situation interaction [$F(1, 76) = 5.79$, $p = .02$].[2] The interaction results from the much smaller effect of social preferences in the public condition, compared to the private condition (see Figure 1). In the public condition, altruists and egoists gave an average of 4.12 and 3.66, respectively, a difference of only .46 resource points, $t(78) = 1.71$, $p = .09$. However, in the private condition, altruists (3.24) gave almost twice as much as egoists (1.72), $t(78) = 3.68$, $p \leq .001$.

While there was a difference in altruists' giving in the public and private conditions (.88, $t(50) = 3.65$, $p = .001$), it was less than half the difference as that of egoists' level of giving

[2] The ordering variable was not significant (p = .54), nor was the three-way interaction between ordering, situation, and person (p = .78). However, the interaction between person and ordering (p = .05) and the interaction between situation and ordering were both significant (p = .01). The person x ordering effect reflects a tendency for altruists to give more (in the public and private conditions) when the public condition comes first, compared to when it comes second. The situation x ordering condition reflects a greater difference between the public and private conditions when the public condition was second, compared to when it was first. Because both these effects could have resulted from participants being told the amount (1/2 the resource pool) the ostensible gave to them following the participant's decision, we change this aspect of the design for our second experiment. In any case, as shown in the main body of text, our predictions hold net of these interaction terms.

between the two conditions (1.93, t (28) = 7.00, $p \leq .001$). The greater difference in behavior across the two conditions for egoists, compared to altruists, is statistically significant, t (78) = 2.74, p = .008. Thus, altruists' level of giving was much less sensitive to the potential for indirect reciprocity.

Discussion

These results support the predicted *person x situation* interaction. While we observed a main effect for the presence or absence of reputational incentives, this main effect was qualified by the interaction of participants' preferences and whether the situation was anonymous or not. We found no support for the alternative predictions of either 1) an unqualified main effect of social preferences, or 2) an unqualified main effect of situation. Nor did we find support for the insincere altruism hypothesis that suggests the presence of reputational incentives will impact altruists more strongly than egoists. Instead, our results showed that while egoists tended to act prosocially only in public, altruists acted prosocially not only in public, but also in private where no reputational incentives for altruism were present.

Study 2: Replication and Extension to Indirect Reciprocity Behavior

Overview

Study 2 has several goals. First, it was designed to replicate the findings from the first study under different conditions. Most significantly, for Study 2 participants there was a time-lag of two- to four-weeks between completion of the social value orientation measure and the behavioral component of the study. As discussed later, this allowed us to ensure that the social values measure did not somehow prime participants' behaviors.

More substantively, whereas Study 1 only addressed altruists' versus egoists' responses to the presence or absence of an opportunity to benefit from future indirect reciprocation, Study 2

also explores the indirect reciprocity of egoists and altruists. As noted earlier, we already know from previous work that people *do* indirectly reciprocate prosocial behaviors directed at others (Milinski et al. 2002a). Our second study explores how indirect reciprocity varies by whether the other's prosocial behavior occurred in the presence or absence of reputational incentives. We also investigate how indirect reciprocity may vary by whether the indirect reciprocator is an egoist or altruist. We address each of these issues in turn.

Our first study created conditions under which prosocial behaviors could occur in public or private. This was a necessary oversimplification, as private behavior may become public: for instance, there are many cases in which a presumed private act (helping or failure to help) may nonetheless be observed or found out by others. At the most basic level, the person who was helped (or not) may tell others, who may then indirectly reciprocate (when help occurred) or withhold reciprocation (when it did not). We suggest that indirect reciprocators "discount" prosocial behaviors that occur in the presence of reputational incentives (see, e.g., Trivers 1971) relative to prosocial behaviors that were done more anonymously. If so, we should find that people indirectly reciprocate *less* when there are reputational incentives for giving at the first order.

We also address whether altruists reward others' prior giving more than egoists do. Assuming no reputational or strategic returns on indirect reciprocation, altruists are expected to indirectly reciprocate at higher levels than egoists.

Finally, we explore the potential interaction between person and situation in indirect reciprocation. Specifically, we propose that, compared to altruists, egoists' indirect reciprocation will show a greater sensitivity to the presence or absence of reputational incentives in others' giving. There are a number of reasons why we expect such a pattern. The first stems directly

from the literature on projection (Iedema and Poppe 1994; Orbell and Dawes 1991, 1993). According to this line of reasoning, in trying to determine the sincerity of others' giving in the presence or absence of reputational incentives, altruists and egoists will project their own behavioral propensities onto others. Thus, altruists will assume that giving behavior is sincere in public or private settings, while egoists will perceive giving as sincere only in private settings.

Another line of reasoning that would lead to the same prediction is that egoists simply attend more closely to the presence or absence of reputational incentives, both in prosocial behavior and in indirect reciprocation of others' prosocial behavior, while altruists do not attend to reputation, behaving prosocially whether others will know or not and rewarding others' prosocial behavior regardless of whether it was done anonymously or publicly. The goal of our second study is not to demonstrate which, if either, of these slightly different mechanisms guides an interaction effect; rather we simply investigate if there is such an interaction.

Summing up, in addition to providing an additional test of our person x situation hypothesis presented earlier, Study 2 is designed to test three predictions about indirect reciprocation:

Hypothesis 2a: Altruists will indirectly reciprocate at higher levels than egoists.

Hypothesis 2b: Individuals will discount prosociality that occurs in the presence of reputational incentives. As a result, they will give higher levels of indirect reciprocation for "private" compared to "public" prosocial actions.

Hypothesis 2c: Egoists will respond more strongly than altruists to the presence versus absence of reputational incentives in others' prosocial behaviors: Egoists will be more likely to indirectly reciprocate prosocial acts that occur in the absence of reputational incentives than those that occur in the presence of reputational rewards. This difference should be smaller, or non-existent, for altruists.

Methods

Design

The second study was conducted using the same population of participants and physical location as Study 1. A total of 70 students (58.6% female), ages 18-27 participated in exchange for partial course credit and the opportunity to earn up to $150 through two $75 raffles. There was one between-subject factor (egoist or altruist) and two within-subject factors. The first within-subjects factor is identical to the one from Study 1: whether the participant's decision in the dictator game was public or private. We measured first order prosocial behavior in each of these decisions. The second within-subject factor was whether an ostensible other's decision in the dictator game was public or private. We measured indirect reciprocity in each of these decisions. Except where noted below, the procedures were like those used in Study 1.

Procedure

As part of course credit, students in two large introductory sociology classes completed a questionnaire containing the social values measure used in Study 1. Two to four weeks later, the students were scheduled in groups of ten to twelve for a study of social decision-making. Upon entering the laboratory, each participant was escorted to a private subject room, where they read and completed a consent form, and then read detailed instructions. Sixty of the seventy participants made at least six choices consistent with a given social value orientation (25 altruists, 35 egoists).

Prosocial Behavior Opportunities: After completing instructions, as in Study 1, participants made choices in a "public" and "private" condition of a dictator game, each with a different other. Whether the participant first made a decision in the public or private condition was

determined randomly. For each of these decisions, the participant decided how many of ten resources (each worth one raffle ticket for two $75 raffles) to transfer to the dependent other and how many to keep for himself or herself. In the public condition, the participant was told that his or her decision would be relayed to a third person who would subsequently decide how much of a 16 point resource pool to transfer to the participant, and how much to keep for himself or herself. (In order to reduce the chances that ostensible others' choices would affect participants' behaviors, we did not provide participants with feedback about the amounts others decided to send to them. As shown below, this change in protocol eliminated the ordering effect observed in Study 1.) Thus, except for the resource pool amount the participants divided (ten versus eight units), the resource pool amount the other divided (18 versus 12), the resource type (raffle tickets versus dollars), and the lack of feedback on others' choices, this stage of the procedure was very similar to its counterpart in Study 1.

We had two objectives in replicating the above procedures. First, we wanted to assess the robustness of the Study 1 results under slightly different conditions (e.g., with spacing between the social value orientation measure and behaviors). But we also wanted to expose participants to the public and private conditions so that they would understand the different types of incentives (i.e., prosocial versus reputational) that exist in each condition. This, we reasoned, would allow participants to make informed attributions about the possible motives for others' behaviors and, in turn, affect participants' tendencies to indirectly reciprocate.

Indirect Reciprocity Opportunities

After the two first order prosocial behavior opportunities, each participant made two indirect reciprocation choices. (In all cases, instructions again emphasized that they would never interact with any given other more than once.) The two conditions were analogous to the public and

private conditions of the prosocial behavior decisions. For each indirect reciprocity decision, participants were informed how much of a twelve point resource pool one ostensible participant (the "Decider") sent to another (the "Receiver"). The instructions for both indirect reciprocity decisions started as follows:

> Now you will be paired with a different participant, [P]. You will be the Decider and [P] will be the Receiver. You have 16 resource points, each worth one $75 raffle ticket. Your task is to decide how many (if any) of the sixteen raffle tickets to transfer to [P]. [P] was paired with a different participant [H] in an earlier interaction. In that interaction, [P] transferred *6* out of *12* raffle tickets to [H] …

Thus, for control purposes, and ease of comparability across conditions, in both conditions, participants were told that [P] sent exactly half (six) of the twelve point resource pool to [H].

The remainder of the instructions differed by condition. For the case in which the ostensible other made his or her decision in public, the instructions continued:

> …[P] was told that this information would be passed on to you (and the recipient [H]) before making his or her choice, and that you would be making a decision about how much of a pool of sixteen points to allocate to him/her.

Thus, the instructions for the public condition clearly indicated the presence of reputational incentives for [P's] behavior.

For the condition in which [P] made his or her decision in the absence of reputational incentives, the instructions continued:

> … [P] was not told that this information would be passed on to you or any other participants (other than the recipient [H]) before making his or her choice. Nor was [P] told that you would be making a decision about how much to allocate to him/her in this phase of the study.

For both conditions, the instructions further noted that the participant's decision (about how much to transfer to [P]) would not be relayed to any other participants, and that he or she would not be paired with [P] or [H] at any other point during the study.

After participants made all four decisions (the two 1st order decisions, and the two 2nd order decisions), they were given lottery tickets, thoroughly debriefed, and checked for suspicion. The entire procedure took about 45 minutes.

Results

Suspicion Check

Participants reported high levels of suspicion in the final decision scenario (the second indirect reciprocity decision) based on the improbability that both prior participants had given exactly half of the resource pool to the dependent other. Because our cover story for the final decision scenario was ineffective, we focus only on the information we collected before these suspicions occurred, i.e., the two prosocial behavior decisions and the first of the two indirect reciprocity decisions.[3]

Replication of Study 1 Results

Descriptive Statistics. The overall results for the prosocial behavior decisions are highly consistent with those of Experiment 1. The modal response for both egoists and altruists in the public condition was to give half of the resource pool. But, as in the first experiment, important differences emerge between altruists and egoists in the private condition. The modal response of altruists was to give half in the private condition, whereas 77% of egoists gave less than half. Figure 2 displays the mean proportions given for each condition.

[Figure 2 about here]

[3] Evidence of the confounding effects of suspicion on the responses in the final decision can be seen by comparing indirect reciprocation amounts in the private condition when it was the first indirect reciprocity decision ($M = 7.19$) and thus suspicion would not have been a problem, to when it was the last indirect reciprocity decision ($M = 5.22$, $p \leq .05$) when suspicion was higher.

Once again, a repeated measures ANOVA shows a significant main effect for person [F (1,56) = 13.22, p ≤ .001], situation [F (1,56) = 14.24, p ≤ .001], and the person x situation interaction [F (1,56) = 4.24, p ≤ .05]. Unlike Experiment 1, neither the main effect of ordering, nor any of the interaction effects were significant (all Fs ≤ 1.28). This suggests that the ordering effects observed in Study 1 likely stemmed from feedback about ostensible others' behaviors.

As in Experiment 1, the interaction results from the smaller effect of social value orientation in the public condition, compared to the private condition (see Figure 2). In the public condition, altruists and egoists gave an average of 4.80 and 4.11, respectively, a difference of only .69 resource points, t (58) = 1.44, p = .16. However, in the private condition, altruists (4.32) gave significantly more than egoists (2.46), t (58) = 4.51, p ≤ .001.

There was no significant difference in altruists' giving in the public and private conditions (t (48) = 1.34, p = .19). However, there was a large difference between egoists' level of giving between the two conditions (t (68) = 3.56, p ≤ .001). The greater difference in behavior across the two conditions for egoists, compared to altruists, is statistically significant, t (58) = 2.08, p ≤ .05. Thus, as in Experiment 1, we found that altruists' level of giving was much less sensitive to the potential for indirect reciprocity.

Indirect Reciprocity Results

Figure 3 displays the average proportion of resource pools given by altruists and egoists in the indirect reciprocity stage of Experiment 2. An ANOVA shows a significant main effect for person [F (1,56) = 6.78, p ≤ .01] and situation [F (1,56) = 4.05, p ≤ .05]. Although the omnibus F statistic for the person x situation interaction failed to reach significance [F (1,56) = .76, p = .39], planned comparisons outlined below show tentative support for Hypothesis 2c. We now turn to

the hypotheses outlined earlier about the role of heterogeneous social preferences and the presence of ulterior motives in situations on indirect reciprocity.

[Figure 3 about here]

Beginning with the main effect of person: In line with Hypothesis 2a, altruists gave more via indirect reciprocity (7.60) than did egoists (6.17), $t(58) = 2.58$, $p \leq .01$. Thus, altruists' prosocial behavior extends beyond first-order prosocial behavior; they also indirectly reciprocate at higher levels than do egoists.

Turning to the main effects of situation (Hypothesis 2b), as shown in Figure 3, participants engaged in less indirect reciprocation when the other's prosociality occurred in public (6.11) versus private (7.33), $t(58) = 1.96$, $p \leq .05$. Note, however, that Hypothesis 2c states that this main effect should be qualified by an interaction. Specifically, the hypothesis predicts that, compared to egoists, altruists will show less behavioral variation in the extent to which they indirectly reciprocate prosocial behaviors that occurred in the presence versus absence of reputational incentives.

Does the presence or absence of reputational incentives for the other's behavior have a greater impact on egoists' level of indirect reciprocity, as suggested by Hypothesis 2c? Compared to altruists' indirect reciprocity in the public versus private conditions (Mean difference = .67, p = .30), egoists showed moderately more variation in indirect reciprocation for public versus private prosocial acts (Mean difference = 1.70, p = .06). Thus, these findings parallel those from the first order prosocial behavior data: egoists respond more strongly than altruists to reputational incentives not only for their own prosocial behaviors, but also the prosocial behaviors of others.

Discussion

The results from the prosocial behavior decisions of Study 2 replicate the results of Study 1 and provide additional support for our hypothesis describing prosocial behavior as a product of a person *x* situation interaction. Furthermore, even with reduced statistical power (due to our inability to use the final indirect reciprocity decision), we found support for all three indirect reciprocity hypotheses.

First, the Study 2 extensions found that not only do altruists show greater prosocial behavior at the first order, but also show greater tendencies to indirectly reciprocate others' prosocial behaviors. We also found that individuals tend to discount altruistic behavior that occurs in the presence of reputational incentives, with egoists doing this significantly more so than altruists. This last pattern implies that, to the extent that indirect reciprocators find out about altruists' higher levels of giving in private settings, this greater altruism can pay off through higher indirect rewards for true prosocial behavior. The irony is that the greatest reputational rewards may go to those who give in situations with little to no apparent reputational incentive, which are nonetheless found out by others somehow. Yet the individuals who are most likely to give in the absence of incentives (altruists) are those who are presumably not pursuing reputational gains.

General Discussion

Our results contain a number of important take-home messages for social scientists interested in problems of altruism, prosocial behavior, and reputations. Ideally, the results of the present research would be used to inform theoretical models of the conditions under which indirect reciprocity systems create niches for various social values. We believe that such models would be beneficial to both social values and indirect reciprocity research. On the one hand, social values researchers have demonstrated an array of important *effects* of social values, but

have devoted little attention to explaining why various social values exist (for an exception, see Van Lange et al. 1997). On the other hand, like theories of kin selection and reciprocal-altruism, indirect reciprocity perspectives have assumed a homogeneous view of the actor, one guided by self-interest.

The question thus becomes: How might systems of indirect reciprocity support both egoism and altruism? Clearly, if altruists are less affected by the presence of onlookers than are egoists (as shown in both of our studies), egoists will save the cost of prosocial behavior when no one is present. Why then wouldn't egoism necessarily squeeze out altruism in evolutionary models?

Results from our second study suggest one possible answer. Study 2 showed that egoists gave higher rewards to those who helped in the absence of reputational incentives than they did to those who gave in public settings, i.e., they rewarded behavior typical of altruists. We argued that this is because egoists are more likely to discount the prosocial actions of others when clear ulterior motives exist. Because altruists are more likely than egoists to help in the absence of conspicuous bystanders, it follows that they will be more highly rewarded in those situations where their generosity is detected. The pattern of indirect reciprocity behavior exhibited by egoists, in particular, favors the prosocial behavior of altruists - who give in the absence of ulterior motives - more than it does other egoists. This finding may help provide a proximal, though incomplete, explanation of how egoistic and altruistic preferences could be simultaneously stable in the population, since altruists' losses relative to egoists in private settings may be recouped through greater indirect reciprocity benefits than egoists when these behaviors are detected by third parties.

While our research focuses on *rewards* via indirect reciprocity, we might reasonably expect parallel discrepancies to emerge for punishments. Recent research shows that third party

punishment plays an important role in generating prosocial behaviors (e.g., Fehr and Fischbacher 2004). Future research should address, for instance, whether egoists, because of their lower tendencies to help in ostensibly private situations, are the targets of greater punishment from onlookers.

Related research points to other potential costs of egoism, including the detectability of insincerity (Frank 1988). But these few examples are sufficient to illustrate our point: a number of factors relevant to indirect reciprocity processes may allow altruism to overcome the selection forces working against it. Hopefully, our experimental results will help motivate formal models that incorporate these and related factors.

In addition to the literatures on altruism and indirect reciprocity, our findings speak to the growing focus on reputations in the sociology of culture (Espeland and Sauder 2007; Fine 2001). As Fine (2001: 24) notes, the sociological study of reputations has traditionally focused on how reputations serve the interests of society. But recent work takes a more micro-analytic approach, addressing how individuals (as actors or perceivers) actively construct and reconstruct reputations for themselves and others. This shift in focus to a micro level highlights a number of means through which reputations can be contested. For instance, because of the behavioral inconsistency of egoists (across public and private opportunities for helping) and because private failures to help may become public, we should expect that egoists will be more likely than altruists to hold contested reputations, or reputations that are "in play" (see Fine 2001: 10*ff*).

In addition, the differing attributions altruists and egoists apparently make about others' prosocial actions shows one mechanism through an identical reputation-relevant behavior may be contested. Questions for future research include how these competing attributions are relayed to others (e.g., via gossip) and which types of attributions "stick" to help create and solidify

reputations. Addressing these types of questions in future work could lead to a better understanding of the dynamic nature of reputations in indirect reciprocity systems, and more fully link the sociological literature on reputations with the literature on indirect reciprocity.

Conclusion

The ubiquity of prosocial behavior among humans has long been a key puzzle in the social sciences. While traditional approaches, such as kin-selection or reciprocal altruism, yield a number of insights, they cannot explain the diversity and degree of prosociality we observe. As a result, a host of new theories and empirical studies of indirect reciprocity have emerged to help explain prosociality. In the bulk of these models, indirect reciprocators act prosocially in order to gain reputational benefits. Thus, the potential for prosociality hinges on whether the act may be observed by others who will then bestow benefits on the helper at some later point in time.

We have argued here that existing models of indirect reciprocity are incomplete for exactly the same reason that models of kin selection and reciprocal altruism are: they rest on a homogeneous view of the actor that belies observed heterogeneity in past research. These models yield misleading conclusions because they predict that prosocial behavior does not happen when reputational incentives are absent. Our work addresses this limitation in the indirect reciprocity literature by incorporating models of heterogeneous social preferences.

At the confluence of research on indirect reciprocity and models of heterogeneous social preferences is the prediction that prosocial behaviors depend on an interaction of the person and situation. The present research shows that egoists respond in predictable ways to the presence or absence of reputational incentives. Helping behavior of altruists, on the other hand, is much less affected by opportunities to garner reputational incentives. Instead, altruists act altruistically at high levels not only when reputational incentives exist, but also when they do not. We found

parallel patterns for indirect reciprocity. For instance, compared to altruists, egoistic indirect reciprocators respond more strongly to whether others' prosocial behaviors may have been motivated by reputational concerns.

Perhaps most significantly, our findings show the sincerity of some actors' altruistic behaviors, since participants classified as altruistic behaved prosocially even in the absence of reputational incentives. Importantly, however, our work also shows that *situations* determine the visibility of individuals' underlying dispositions towards altruism. Dispositions that produce highly divergent behaviors when no conspicuous bystanders are present may result in nearly indistinguishable public behaviors. The regrettable paradox is that we may only be able to observe others' underlying nature in precisely those situations when it is typically unobservable.

References

Alexander, Richard D. 1987. *The Biology of Moral Systems*. New York: De Gruyter.

Batson, C. Daniel. 1987. "Prosocial Motivation: Is it ever Truly Altruistic?" *Advances in Experimental Social Psychology* 20:65-115.

Bem, Daryl and Charles Lord. 1979. "Template matching: A Proposal for Probing the Ecological Validity of Experimental Settings in Social Psychology." *Journal of Personality and Social Psychology* 37:833-46.

Bowles, Samuel and Herbert Gintis. 2004. "The Evolution of Strong Reciprocity: Cooperation in Heterogeneous Populations." *Theoretical Population Biology* 65:17-28.

De Cremer, David and Paul Van Lange. 2001. "Why Prosocials Exhibit Greater Cooperation than Proselfs: The Roles of Social Responsibility and Reciprocity." *European Journal of Personality* 15:5-18.

Dugatkin, Lee Alan and David Sloan Wilson. 1991. "A Strategy for Exploiting Cooperators in a Patchy Environment." *The American Naturalist* 138:687-701.

Espeland, Wendy Nelson and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." 113:1-40.

Fehr, Ernst and Urs Fischbacher. 2004. "Third-party Punishment and Social Norms." *Evolution and Human Behavior* 25:63-87.

_____. 2005. "Human Altruism - Proximate Patterns and Evolutionary Origins." *Analyse and Kritik* 27:7-47.

Fine, Gary Alan. 2001. *Difficult Reputations: Collective Memories of the Evil, Inept, and Controversial.* University of Chicago Press.

Frank, Robert H. 1988. *Passions within Reason.* New York: W.W. Norton.

Ghiselin, Michael. 1974. *The Economy of Nature and the Evolution of Sex.* Berkeley: University of California Press.

Hamilton, William D. 1964. "The Genetical Evolution of Social Behaviour. I and II." *Journal of Theoretical Biology* 7:1-52.

Iedema, Jurjen and Matthijs Poppe. 1994. "Effects of Social Value Orientation on Expecting and Learning Others' Orientations." *European Journal of Social Psychology* 24:565-69.

Kurzban, Robert and Daniel Houser. 2005. "Experiments Investigating Cooperative Types in Humans: A Complement to Evolutionary Theory and Simulations." *Proceedings of the National Academy of Sciences* 102:1803-07.

Leimar, Olof and Peter Hammerstein. 2001. "Evolution of Cooperation Through Indirect Reciprocity." *Proceedings of the Royal Society B* 268: 745-53.

Liebrand, Wim B.G. 1986. "The Ubiquity of Social Values in Social Dilemmas." Pp. 113-133 in *Experimental Studies of Social Dilemmas*, edited by Henk Wilke, Christel Rutte, and David Messick. Frankfurt: Peter Lang Publishing Company.

Liebrand, Wim B.G., Henk A.M. Wilke, and Fred J.M. Wolters. 1986. "Value Orientation and Conformity: A Study Using Three Types of Social Dilemma Games." *Journal of Conflict Resolution* 30:77-97.

Milinski, Manfred, Dirk Semmann, and Hans-Jurgen Krambeck. 2002a. "Reputation Helps Solve the 'Tragedy of the Commons.'" *Nature* 415:424-26.

_____. 2002b. "Donors to Charity Gain in Both Indirect Reciprocity and Political Reputation." *Proceedings of the Royal Society of London* 269:881-83.

Nowak, Martin. 2006. "Five Rules for the Evolution of Cooperation." *Science* 314:1560-63.

Nowak, Martin and Karl Sigmund. 1998. "Evolution of Indirect Reciprocity through Image Scoring." *Nature* 393:573-77.

Orbell, John M. and Robyn M. Dawes. 1991. "A 'Cognitive Miser' Theory of Cooperators' Advantage." *American Political Science Review* 85:515-28.

_____. 1993. "Social Welfare, Cooperators' Advantage, and the Option of Not Playing the Game." *American Sociological Review* 58:787-800.

Panchanathan, Karthik and Robert Boyd. 2004. "Indirect Reciprocity can Stabilize Cooperation without the Second-order Free Rider Problem." *Nature* 432:499-502.

Rockenbach, Bettia and Manfred Milinski. 2006. "The Efficient Interaction of Indirect Reciprocity and Costly Punishment." *Nature* 444:718-23.

Semmann, Dirk, Hans-Jurgen Krambeck, and Manfred Milinski. 2004. "Strategic Investment in Reputation." *Behavioral Ecology and Sociobiology* 56:248-52.

Simon, Herbert A. 1990. "A Mechanism for Social Selection and Successful Altruism." *Science* 250:1665-68.

Smith, Eric A., and Rebecca A. Bliege Bird. 2000. "Turtle Hunting and Tombstone Opening: Public Generosity as Costly Signaling." *Evolution and Human Behavior* 21:245-61.

Sugden, Robert. 1986. *The Economics of Rights, Cooperation, and Welfare*. Blackwell: Oxford.

Sober, Elliott, and David Sloan Wilson. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.

Trivers, Robert. 1971. "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology* 46:35-57.

Van Lange, Paul A.M. 1999. "The Pursuit of Joint Outcomes and Equality in Outcomes: An Integrative Model of Social Value Orientation." *Journal of Personality and Social Psychology* 77:337-49.

Van Lange, Paul A.M, Wilma Otten, Ellen M.N. De Bruin, and Jeffrey Joireman. 1997. "Development of Prosocial, Individualistic, and Competitive Orientations: Theory and Preliminary Evidence." *Journal of Personality and Social Psychology* 73:733-46.

Wedekind, Claus and Manfred Milinski. 2000. "Cooperation Through Image Scoring in Humans." *Science* 288:850-2.

Zahavi, Amotz. 1995. "Altruism as a Handicap - the Limitations of Kin Selection and Reciprocity." *Journal of Avian Biology* 26:1-3.

Brent Simpson is Associate Professor of Sociology at the University of South Carolina. His recent publications have addressed regional differences in trust in the United States (*Social Forces* 2006), how generalized trust influences the use of social relations for economic exchanges (*Social Networks* 2007, with Tucker McGrimmon), and race-category differences in trust (*Social Forces* 2007, with Tucker McGrimmon and Kyle Irwin).

Robb Willer is Assistant Professor of Sociology at University of California, Berkeley. His works focuses on the microfoundations of social order (e.g. collective action, altruism, trust) and the social psychology of political ideology. His research has appeared in *American Journal of Sociology*, *Annual Review of Sociology*, and *Proceedings of the Royal Society B: Biological Sciences*.