

When It Pays to Be Kind:

The Allocation of Indirect Reciprocity within Power Hierarchies

Abstract

Indirect reciprocity – the notion that third-party observers offer rewards to prosocial actors – is known to increase levels of cooperative behavior. Yet we know relatively little about how people decide to *grant* indirect reciprocity. This process is complex because it relies on assessing moral character, which is unobservable. In the current research, we identify a salient cue in the social environment that observers use to calibrate their indirect reciprocity: power differences. Across three experiments, observers were less likely to offer indirect reciprocity to employees who targeted their generosity toward higher- rather than lower-power co-workers. Indirect reciprocity was measured through the allocation of participants' own financial resources (Experiments 1 and 2), as well as behavioral intentions (Experiment 3). Experiment 3 also showed that this effect is driven by observers' perceptions of actors' motives, which inform assessments of moral character.

KEYWORDS: indirect reciprocity; prosocial behavior; power; perceived motive; moral character

Indirect reciprocity is a driving force in the evolution of human cooperation (Alexander, 1987; Nowak & Sigmund, 2005). It describes the process by which people offer more resources to individuals who help *others*. Indirect reciprocity is at play, for example, when an employee offers help to a generous co-worker, even if the employee has never personally benefited from the co-worker's generosity in the past. Through acts of kindness and generosity (i.e., prosocial behavior), prosocial actors signal to others that they are good people, possessing high moral character. These attributes are valued in social groups, so others tend to reward them with higher levels of indirect reciprocity (Nowak & Sigmund, 2005; see also van Vugt, Roberts, & Hardy, 2007). Critically, this means that a prosocial act can ultimately yield benefits that exceed the initial costs to the actor (Wedekind & Braithwaite, 2002). The possibility of eventual benefits has been used to explain why people often choose to cooperate, absent immediate incentives (Nowak & Sigmund, 2005).

Research to date has shed considerable light on this dynamic from the prosocial actor's perspective: actors are more likely to act in altruistic ways when third-party observers will know about their actions and can offer indirect reciprocity as a reward (Milinski, Semmann, & Krambeck, 2002; Wedekind & Milinski, 2000). For example, employees are more likely to engage in prosocial organizational behavior when they expect recognition and rewards (McNeely & Meglino, 1994). In marked contrast, however, considerably less research has been conducted on indirect reciprocity from the observer's perspective to understand when and why observers *offer* indirect reciprocity to prosocial actors. This is surprising, since an observer's decision to reward a prosocial act is as much an element of the indirect reciprocity process as the actor's decision to act prosocially in the first place.

In offering indirect reciprocity, observers seek to reward people who have high moral character (Nowak & Sigmund, 2005). This is a challenging task, however, because moral character is unobservable. Making it even more difficult, people often act in generous ways,

but do so for selfish reasons (Ariely, Bracha, & Meier, 2009; Bénabou & Tirole, 2006). So how do observers navigate this difficult task? Our goal in the current research is to identify a signal that observers use to calibrate their indirect reciprocity allocations.

We look to the social context as a rich source of information for observers. In particular, we focus on a basic feature of all social interactions that is quickly noticed and encoded by observers (Zitek & Tiedens, 2012): power differences. We predict that observers use information about power differences between prosocial actors and their targets to infer the selfless versus selfish nature of the actor's motive when performing the prosocial act. This in turn informs observers' assessments of actors' moral character, and ultimately, determines their allocations of indirect reciprocity.

Our work makes two important contributions. First, we reveal how indirect reciprocity is allocated in contexts in which observers witness a single prosocial act, and have little to no private or historical information about the actor. We demonstrate that observers use features of the social context as a valuable guide to infer motive and assess moral character. Second, by identifying power as a signal that affects indirect reciprocity, our findings apply to prosocial acts in which power differences between givers and receivers are salient. Power differences are both highly visible and constitute a fundamental feature of social perceptions. This means that, for a large portion of prosocial acts, observers can use information about power differences to make decisions about how to allocate indirect reciprocity.

Assessing Motive and Moral Character

Research to date has repeatedly demonstrated that individuals who signal having higher moral character are offered more indirect reciprocity from observers (Hoffman, Yoeli, & Nowak, 2015; Jordan, Hoffman, Nowak, & Rand, 2016; Simpson & Willer, 2008). This pattern belies a functional dynamic in which those who display traits that are valuable in

social groups – kindness, generosity, cooperation – gain greater resources and status (Leimar & Hammerstein, 2001; Nowak & Sigmund, 1998; Zahavi & Zahavi, 1997). Moral character is unobservable, however, so observers can only assess it using visible signals. What are these signals?

Actions are a good starting point because they are highly visible. An actor who performs generous prosocial acts communicates to observers that he or she has high moral character, an effect that has been shown empirically (Funder, 2004; Gilbert, 1989). Observing multiple acts over time may help establish a pattern of behavior consistent with higher versus lower moral character (e.g., Wedekind & Milinski, 2000). However, observers often have no insight into a person's behavior over time; they must often decide how much indirect reciprocity to award after observing only a single act.

For a single prosocial act, a helpful indicator of moral character is the motive driving it. Why did a person donate money? Why did they help out a colleague? And in particular, was it driven by a selfless or selfish motive? This assessment of an actor's motive has emerged across the literature on prosocial behaviour as a basic dimension on which such acts are judged (e.g., Fein, 1996; Miller, 1999). Importantly, this attribution affects assessments of moral character: individuals whose prosocial acts are seen to be driven by more selfish (i.e., less altruistic) motives are assessed as having lower moral character. In one study for example, participants were told that a person gave blood either to earn money or to help people. When they believed the donor gave blood to earn money, they judged this person as having lower moral character (Carlson & Zaki, 2018).

Further, observers' attribution processes follow a predictable pattern of discounting (Crocker, 1981; Kelley, 1972). Once a selfish motive emerges as a potential driver of a prosocial act, observers are less likely to believe it was driven by *selfless* motives (Inesi, Gruenfeld, & Galinsky, 2012). Therefore, even if there is much evidence to suggest an

altruistic intent, any indication of selfish gain leads to an overall perception of reduced altruism (Critcher & Dunning, 2011), and to lower moral character assessments (Carlson & Zaki, 2018).

Like with moral character, an observer cannot know a prosocial actor's true motive. However, visible signals of motive can leak out, such as a personal connection to the charity (Lin-Healy & Small, 2012), or a romantic interest volunteering at the same event (Newman & Cain, 2014). Each of these signals suggests that a prosocial act may be partly driven by selfish motives, which in turn affect assessments of moral character (Reeder, Kumar, Hesson-McInnis, & Trafimow, 2002). For example, when a person brags about donating money to a charity, observers read this as a signal of a selfish motive, and therefore believe the donor has lower moral character (Berman, Levine, Barasch, & Small, 2015).

For the purposes of our research, the key question is whether cues about motive ultimately affect resource allocations, i.e., indirect reciprocity. There is some initial evidence to support this relationship. For example, participants offered less indirect reciprocity when they learned that a prosocial actor sought additional information about payoffs before cooperating, or took a longer time to decide (Jordan et al., 2016; see also Simpson & Willer, 2008). These authors theorized that this pattern was driven by perceptions of selfish motives, although this was not tested empirically. Importantly though, these signals are typically hidden from observers. In naturalistic settings, observers don't know how long it took for a person to decide to act generously, how much information they sought, or whether the actor knew their decision would be made public. Therefore, they must rely on more visible cues to make these assessments. What are these signals?

Visible Signals of Motive

We designed the present research to answer this question. Our conceptual starting-point is the notion that observers have access to a host of accessible social information about

the prosocial actor. Social information, like demographic characteristics or social roles, is often visible to observers. More importantly, it has been shown to affect observers' attribution processes (e.g., Deaux & LaFrance, 1998; Pettigrew, 1979; Sekaquaptewa & Espinoza, 2004) and assessments of character (Rudman & Glick, 1999). Accordingly, we suggest that observers use features of the social context surrounding a prosocial act as signals of information about the motive driving the prosocial act, which then informs the observers' assessment of the actors' moral character, and in turn, indirect reciprocity allocations.

Arguably, one of the most ubiquitous social forces is power. Power hierarchies are a fundamental feature of both social relations as well as social perceptions; they are highly visible and are also a key determinant of human behavior (Emerson, 1962; Fast, Gruenfeld, Sivanathan, & Galinsky, 2009; Magee & Galinsky, 2008; Galinsky, Rus & Lammers, 2011; Whitson et al., 2013). Observers are often cognizant of the impact power has on others' behavior and use it to form attributions. For example, when hiring managers accept referrals from higher- rather than lower-power referrers, observers judge them as more selfishly and instrumentally motivated (Derfler-Rozin, Baker, & Gino, 2018). Also, observers are more likely to attribute low-power individuals' actions to the situation, and high-power individuals' actions to their personal preferences (Overbeck, Tiedens, & Brion, 2006).

Here, we suggest that observers use information about power hierarchies as a signal to calibrate their indirect reciprocity. We believe this occurs because power differences signal how selfish versus selfless the motive driving the prosocial act is. Prosocial acts targeted at higher-power individuals are perceived by observers as less altruistically motivated, and therefore indicative of lower moral character. Ultimately, we predict, this drives indirect reciprocity allocations.

Prior theoretical and empirical work is suggestive of this possibility. Jones (1964) noted that power differences affect the motives underlying compliments and praise. Higher-

power individuals have, by definition, greater access and control over desired resources (Emerson, 1962). This, therefore, represents an extrinsic, selfish motive behind kind acts directed at power-holders: people want access to the valued resources and therefore offer kind acts to power-holders as a way of ingratiating and gaining access to them. In line with this notion, studies have shown that recipients of a favor were less likely to think it was selflessly motivated when it came from a subordinate compared to a peer (Inesi et al., 2012; see also Kunstman, Fitzpatrick, & Smith, 2018). Although suggestive, these findings only document the experience of direct beneficiaries (recipients of a favor), who would likely be keenly aware of the power difference between themselves and the prosocial actor. They offer little insight into how power differences between a prosocial actor and their target would affect *observers'* attributions. In line with recent work on power and attributions in the context of advice-taking (Derfler-Rozin et al., 2018), we predict that observers will attribute less altruistic motives to prosocial acts directed at higher-power individuals.

The key question for the purposes of our research is whether power differences, and the related perceptions of motive and moral character, will ultimately predict indirect reciprocity. Research has shown that people have stable preferences for how cooperative and generous they want to be, falling back on personal preferences when allocating indirect reciprocity (e.g., Bolton, Brandts, & Ockenfels, 1998; Evans & van de Calseyde, 2017; Simpson & Willer, 2008). Therefore, independent of context, motive, or moral character assessment, observers may still offer the same level of indirect reciprocity to all prosocial actors. Our goal here is to demonstrate that power differences do indeed act as a signal that affect not only observers' perceptions, but also the amount of resources they are willing to allocate.

In sum, we predict that observers will be less likely to attribute prosocial acts to selfless motives when they are targeted at higher-power (versus lower-power) individuals.

Observers will then use this information to generate an understanding of the actor's moral character, such that prosocial actors who target higher-power individuals will be seen as possessing lower moral character. Finally, we predict that observers use moral character judgments to allocate indirect reciprocity, such that they will offer less indirect reciprocity to individuals whose prosocial acts are targeted at higher- versus lower-power individuals.

Overview of Studies

We tested our predictions across three pre-registered experiments. In the first two experiments, we tested the core hypothesis that participants will offer fewer real resources (i.e., indirect reciprocity) to another individual who has earlier enacted prosocial behavior towards a higher-power compared to a lower-power target. In the first experiment, we also included a condition in which no prosocial behavior was offered. In this way, we can compare how observers respond to actors who choose *not* to behave prosocially vis-à-vis those who offer generosity up versus down the power hierarchy. In the second experiment, we employed a new, longitudinal design to ensure higher external validity. We also substantially increased the pool of resources from which the indirect reciprocity could be offered. Our goal was to demonstrate that, as participants' endowment increases, they are willing to part with more meaningful amounts of money to award indirect reciprocity, and use power differences as a signal to guide their allocations. Finally, in the third experiment, we sought to document the psychological process that underlies our proposed effect by asking participants, first, to what extent they believed the prosocial act was motivated by selfless versus selfish concerns, and second, to assess the moral character of the prosocial actor. We predicted a two-stage indirect effect of relative power on indirect reciprocity through perceived motive and moral character (serial mediation).

All three studies reported in this manuscript were pre-registered. We report all manipulations, measures, and data exclusions. Sample sizes were determined a priori, and we

did not collect further data after analysis. Informed consent was obtained from all participants and all studies received institutional ethics approval. All study materials have been posted online.¹ Data are available to individuals upon request to the authors.

Experiment 1

We designed Experiment 1 to test the prediction that power differences affect indirect reciprocity allocation decisions. To model existing indirect reciprocity paradigms as closely as possible, we drew on published research in this area (Herne, Lappalainen, & Kestila-Kekkonen, 2013; Jordan et al., 2016; Simpson & Willer, 2008). In this paradigm, participants are placed in the observer role and learn about the behavior of another individual that they did not already know. Then, they are given an opportunity to share resources with this person. In Experiment 1, we used the Dictator Game to measure indirect reciprocity, consistent with other research (Herne et al., 2013; Simpson & Willer, 2008). The rules of the Dictator Game are that one person (the Dictator) is given financial resources that they can keep or share with another person. Any money they do not share is theirs to keep. Thus, the amount offered by the Dictator precisely captures the situation in which a person must part with a valuable resource in order to reward someone else.

A second contribution of Experiment 1 is that it includes a condition in which no prosocial act was offered. This provides a helpful comparison for the other conditions in two ways. First, it adds credence to the assumption that participants are indeed offering more indirect reciprocity to prosocial acts than to non-prosocial acts. Second, it provides a baseline to calculate the discount that upward-directed prosocial acts receive in terms of indirect reciprocity. If the results show that upward-targeted prosocial acts garner less indirect reciprocity than downward-directed ones, it would not be clear how these amounts compare

¹ https://osf.io/h2uf6/?view_only=4f47b9604eed4f9188cd

to the amount of indirect reciprocity a person would receive from third parties absent any act of prosociality. By including this additional control condition, we can quantify the discount in indirect reciprocity that occurs as a result of prosocial behavior given up versus down the power hierarchy.

Method

Experiment 1 was preregistered (<https://aspredicted.org/blind.php?x=t7iv9j>).

Participants. We recruited 600 participants from Prolific Academic. We based this recruitment number on a small/medium Cohen's $d = 0.26$ from an earlier version of the study (which contained higher- and lower-power target conditions similar to this study; see Supplemental Materials 2², Study 1SM). We expected that the omnibus effect across these cells plus the added no-prosocial-behavior condition would yield a comparatively weaker effect. Therefore, we conservatively assumed a weak omnibus effect size of $f = .13$. This indicated a total sample size of 576 was necessary to achieve power of .80 (2-tailed). Our final sample after exclusions was 560 participants (55% women; $M_{\text{age}} = 35.90$; $SD_{\text{age}} = 10.30$). Participants were paid £0.39 (£0.30 plus £0.09 bonus money).

As criteria for participation, we sought participants who were full-time employees and who reside in the United Kingdom. We pre-selected for full-time employees so that they could better envision the organization-based scenario. We selected a single country because we wanted all participants to be familiar with the monetary currency used in the materials. We pre-registered to remove participants with duplicate IP addresses (in an effort to remove bots from our dataset) and those who failed attention checks (they were not allowed to complete the study).

² https://osf.io/h2uf6/?view_only=4f47b9604eed4f9188cdba57256a57b8

Procedure. We used a between-participants design with random assignment to the three cells: relative power of target (higher-power vs. lower-power vs. control (no prosocial behavior)). For the full experimental procedures, see Supplemental Materials 1.³

At the start of the study, participants were informed that they would be working on a resource allocation task with another Prolific worker and that they could earn up to 9 pence (p) in bonus money based on their choices. Participants were informed that a longitudinal study had taken place over the past four weeks on the Prolific platform. They were told that the 3-person team's task had been to create content for a website describing vacation packages. Player B – the content creator – had been supervised by a boss (Player A), and supported by an assistant (Player C). Participants also read that the team members had been connected via a communication platform called ChatPlat (e.g., Brooks & Schweitzer, 2011), so they could send messages and content to each other. Finally, we included information about how much the team members had been paid, to highlight the power differences: the boss had earned £10 for the work and the assistant had earned £3. Player B's pay had been contingent on their boss's evaluation, however (£8 for an excellent assessment, £4 for a satisfactory assessment and £2 for a poor assessment).

Next participants read that during Week 3 of the study, Player B had received £0.50 bonus money from the experimenter. The boss and the assistant had not received any bonus money, nor had they known that Player B received any. Player B had had the choice of either keeping the bonus money, or sharing half of it with either the boss or the assistant. Player B had been told that only the recipient would know that they had received a bonus, and had decided to share it.

³ https://osf.io/h2uf6/?view_only=4f47b9604eed4f9188cdba57256a57b8

Depending on the condition, participants read that Player B had decided not to share the bonus money (*control* condition); or had decided to share it with the boss (*higher-power target* condition) or the assistant (*lower-power target* condition).

The Dictator Game. Next, participants started the Resource Allocation task (actually a Dictator Game) played with Player B. Participants played the role of Dictator. They were endowed with 9p of bonus money and were asked to decide how much they wanted to give to Player B. Participants were told that they could keep whatever they did not give to Player B. The dependent measure of indirect reciprocity was how much money participants sent to Player B (0p-9p).

To diagnose whether participants understood the Dictator Game, we asked participants to confirm how much money Player B would receive. Removing participants who answered incorrectly did not affect our results. Finally, participants reported their age and gender. We asked attention check questions throughout the study to make sure that participants understood the content of the study. If participants failed two successive attention check questions, they were directed out of the study at that point.

Results and Discussion

The omnibus ANOVA showed a significant effect of the relative power of the target on indirect reciprocity, $F(2, 557) = 19.81, p < .001, \eta^2 = .07$. Consistent with our prediction, participants offered less indirect reciprocity when they learned that Player B had shared the bonus money with the higher-power boss ($M = 2.65, SD = 2.39$) versus the lower-power assistant ($M = 3.29, SD = 2.31$), $t(557) = -2.66, p = .008, d = 0.27$. Furthermore, participants offered less indirect reciprocity when Player B did not share the bonus (i.e., did not act prosocially) ($M = 1.80, SD = 2.17$) compared to when they shared the bonus with the boss, $t(557) = -3.62, p < .001, d = 0.37$, or the assistant, $t(557) = 6.25, p < .001, d = 0.64$.

Taken together, these results suggest that the indirect reciprocity awards associated with giving versus not giving are 43% lower when generosity is directed up versus down the power hierarchy. We determined this by first calculating the average added benefit of giving to a lower-power individual versus not giving ($3.29 - 1.8 = 1.49$). Then we calculated the average added benefit of giving to a higher-power individual versus not giving ($2.65 - 1.80 = .85$). The discount applied to upward-directed versus downward-directed can be calculated as $(1 - .85/1.49) = .43$, or 43%.

Experiment 2

A core tenet of our predictions is that observers use power differences as a signal to guide their allocation of valuable resources. Experiment 1 offered initial support for this prediction. At the same time, the absolute difference in resources offered to prosocial actors who gave to higher- versus lower-power targets was less than 1p. Although statistically significant, its practical significance in terms of real resources is relatively small. Therefore, in Experiment 2 we increased initial endowment by more than a factor of ten, to £1 (approximately \$1.30 at the time of the study). With this larger endowment, two possible patterns may emerge. One possibility is that participants would continue to offer only trivial amounts to the prosocial actors, and therefore leave the difference in indirect reciprocity between the two conditions as more of a social signal rather than a meaningful difference in resources. Another possibility is that, with a larger endowment, participants would offer larger amounts of indirect reciprocity, such that the difference in rewards for prosocial acts targeted toward higher- versus lower-power individuals becomes more meaningful in financial terms.

Although Experiment 1 described the actions of ostensible participants from an earlier longitudinal study, it is possible that participants did not believe that they were real people. To address this in Experiment 2, we developed an elaborate longitudinal design in which

participants observed and responded to daily actions of an ostensible virtual team over the course of a week. Only when they reached Day 4 of a five-day study did they complete the focal dependent measure. We administered suspicion checks at the end to check how believable the procedure was to participants.

Method

Experiment 2 was preregistered (<https://aspredicted.org/blind.php?x=ns76d2>).

Participants. When comparing reactions to prosocial actors helping higher- versus lower-power others, Experiment 1 showed a $d = 0.27$ (80% power, 2-tailed). Because Experiment 2 makes a similar comparison, we calculated our power analysis based on that effect size and therefore aimed to collect data from 434 participants. First, we posted a pre-screening recruitment survey for 850 participants. We over-recruited considerably because of the longitudinal nature of the study (conducted over 4 days), and because we planned to only invite participants who met the following criteria: participants without duplicate IP addresses; those who passed attention checks; and, in an effort to remove bots, those who provided sensical responses to an essay question (all these exclusion criteria were preregistered). We then invited the 573 participants who met these criteria to participate in a 4-day longitudinal study; we further lost participants over the course of the study due to attrition, duplicate IP addresses, and because they failed to pass additional attention checks. Our final sample was $N = 469$ (65.88% women; $M_{\text{age}} = 37.56$; $SD_{\text{age}} = 13.65$). Participants were paid £2 for completing the longitudinal study and additional money based on their decision in the indirect reciprocity task ($M_{\text{bonus paid}} = \text{£}0.73$; $SD = .26$).

Procedure. We used a between-participants design with random assignment to the two conditions: relative power of target (higher-power vs. lower-power). For the full experimental procedures, see Supplemental Materials 1.⁴

Recruitment. We posted on Prolific that we were recruiting participants for a 5-day longitudinal study. The purpose of the study was described as an attempt to better understand how virtual teams work over time, since the Covid-19 pandemic necessitates more virtual work. As part of the study, participants were told that they would be observe a virtual team for 5 days, provide daily feedback, and be paid 20p each day plus an additional 80p payment for completing all 5 days (i.e., £2 total payment). The recruitment survey was posted on a Sunday. Participants were told that if they wanted to participate, they should plan to take a short survey each day between 8 am and 8 pm (Monday through Friday). They were told that they would need to answer 2-3 short questions about the team they would be assigned to observe, that they would need to participate every single day and that participants who missed a day would not be invited to the following day's study.

To bolster the cover story that they would be observing real teams, we simultaneously posted another study on Prolific, which purported to recruit participants for a study on Virtual Teams as a team member. When participants clicked that link, they received a message saying that the study was full (and were unable to sign up for it).

Day 1. We sent the Day 1 survey on Virtual Team Observation to the 573 participants we recruited on Day 0. Participants were informed that they would observe a 3-person hierarchical team over the course of the week whose job is to create content (text) for a website describing holiday [vacation] packages. They were told that the team they would be

⁴ https://osf.io/h2uf6/?view_only=4f47b9604eed4f9188cdba57256a57b8

observing was comprised of a Boss, Content Creator, and Assistant, who had been assigned to those roles based on their prior work experience. They were told that the Content Creator would write 300 words and then send it to the Assistant, who would copy-edit and proofread it. The Boss would read over the content and evaluate the Content Creator's work. Also, the Content Creator would evaluate the Assistant's copy-editing skills. Furthermore, they were told the team would get together at the end of each day to discuss how they might improve their processes and performance.

To reinforce the power structure, we told participants that the Boss's compensation was fixed at £10, whereas the Content Creator's payment would depend on the Boss's evaluation (£8 for an excellent assessment, £6 for a satisfactory assessment and £4 for a poor assessment). Similarly, the Assistant's compensation would depend on the Content Creator's evaluation (£5 for an excellent assessment, £4 for a satisfactory assessment and £3 for a poor assessment).

Last, participants were asked two questions about the team structure to increase the believability of the cover story. We did not analyze these data because they were not relevant to our hypothesis. We administered attention and comprehension checks and excluded participants based on duplicate IP addresses. This left us with 537 participants eligible for Day 2.

Days 2-3. On Days 2-3, we administered surveys to bolster our cover story that participants were observing a real team over the course of a week. These questionnaires were not relevant to our hypotheses and thus we did not analyze responses. However, we were careful not to provide any individual-level information about the team members because participants may have used such information to form an understanding of the virtual team members' moral character. Since moral character assessments drive indirect reciprocity allocations (Baker & Bulkley, 2014), this could have muted the effect of our key

manipulation (relative power of the target of prosocial behavior). Thus, on Day 2, they were presented with aggregate information about the number of hours worked by the virtual team members on the prior day, and were asked questions about their perceptions of the team's effort and performance. On Day 3, they were shown a sample of the team's writing and were asked to respond to two questions about the text's quality. We also administered attention and comprehension checks and excluded participants from duplicate IP addresses. This left us with 518 participants eligible for Day 3 and 480 participants eligible for Day 4.

Day 4 (Final Day). We again described the virtual team members' task and their hierarchy and pay structure. Participants were then told, "*Today we are interested in understanding how team members' interactions with each other affect how you would react to them in an outside content. On the following pages, we will share information about an interaction that occurred yesterday among the team members.*" Participants were informed that the Content Creator had received a £4 bonus from the experimenter. The Boss and the Assistant had not received any bonus money and did not know that Content Creator had received any. The Content Creator had the option of either keeping the bonus money, or sharing half of it (£2) with either the Boss or the Assistant. Only the recipient would know that the Content Creator had received and shared the bonus. Depending on condition, participants read that the Content Creator had decided to share it with the Boss (*higher-power target condition*) or the Assistant (*lower-power target condition*).

We then told participants that they would be participating in a Resource Allocation task with the Content Creator. They were endowed with a £1 bonus – a substantial amount for Prolific compensation – and were asked to decide how much they wanted to give to the Content Creator. Participants were told that they could keep whatever they did not give to the Content Creator. We measured indirect reciprocity in terms of how much money participants

sent to the Content Creator (0p-100p). Participants were informed that the Content Creator would be informed of the allocation, if any, on Day 5.

Again, participants were asked a series of comprehension checks. After exclusions, the final sample was $N = 469$.

Finally, we evaluated participants for suspicion, provided debriefing information, and told them that there would not be a questionnaire on Day 5 and that they would receive the full payment for the study, plus any amount they chose not to allocate to the Content Creator.

Evaluation for Suspicion. To understand whether participants were suspicious that the team members were not real participants, we used a funnel debriefing procedure (e.g., Chartrand, van Baaren, & Bargh, 2006), in which we asked a series of questions meant to probe suspicion. We started with general questions about their beliefs about the purpose of the study before probing whether they thought the other participants were real. In total, 94 participants (20.04%) indicated suspicion.

Results

The effect of condition on indirect reciprocity was significant, $t(449.94) = 5.62, p < .001, d = 0.52$.⁵ Participants offered less indirect reciprocity when they learned that the Content Creator had shared bonus money with the higher-power Boss ($M = 20.49, SD = 23.58$) versus the lower-power Assistant ($M = 33.77, SD = 27.38$).

When we removed participants who reported that they thought the virtual team members were not real participants, results remained robust, ($M_{Boss} = 22.78, SD = 24.41; M_{Assistant} = 35.34, SD = 26.84$), $t(373) = 4.75, p < .001, d = 0.49$.

Discussion

⁵ The data showed a violation of the assumption of homogeneity of variance; we therefore report corrected t -tests.

Experiment 2 replicates and extends the findings of our first study. As before, we found that observers gave more of their own bonus money to other individuals when this person had done something kind for a lower- versus higher-power individual. In Experiment 2, however, this emerged when participants were given a significantly larger endowment, and in a context that closely mirrors how indirect reciprocity unfolds in the real world. That is, participants learned over time about team's interactions. Then, they learned about a specific actor's behavior towards another team member and were given the opportunity to help the actor in a separate context.

Experiment 3

Our theory makes clear predictions about the mechanisms that underlie the results from Experiments 1 and 2. We hypothesized that observers would believe prosocial behaviors targeted at higher-power actors are less altruistically motivated than those targeted at lower-power actors, which in turn would be associated with perceptions that the prosocial actor has lower moral character. Ultimately, this character assessment would be associated with differences in indirect reciprocity allocations. We therefore measured these variables (pre-registered) to test for a serial indirect effect.

A second goal of Experiment 3 was to test the limits of our predictions by de-coupling the beneficiary from the target of the prosocial act. In Experiments 1 and 2, the hierarchical counterpart was also the beneficiary of the prosocial act. That is, the prosocial actor gave money to either their boss or subordinate. Based on the norms of direct reciprocity (Gouldner, 1960), the beneficiary should feel pressure to return the favor. Since by definition higher-power individuals are able to offer higher-value reciprocity than lower-power individuals, then prosocial acts offered to higher-power individuals are judged by observers as more selfishly motivated. The results from Experiments 1 and 2 are consistent with this dynamic, and show that it affects indirect reciprocity allocations.

But what if the boss or subordinate never receives any benefit at all? Instead, what if the prosocial actor helps a needy third-party (e.g., volunteers at a homeless shelter), but in such a way that their generosity is *witnessed* by either their boss or subordinate? Will observers still be more suspicious of prosocial acts merely witnessed by a higher-power target, even though the target received no concrete benefit? We predicted that they would. Research has shown that observers are sensitive to strategic uses of prosocial behavior to build a positive image, and believe such actors have lower moral character (Berman et al., 2015). Since building a positive image with a boss is likely to be more lucrative than building one with a subordinate, we predicted that observers would confer less indirect reciprocity when the prosocial target was higher-power because they believe the individual is less altruistically motivated and has a lower moral character.

A final goal of Experiment 3 was to test our predicted effects when the participant was in the same organization (i.e., social group) as the prosocial actor and the target. In Experiments 1 and 2, the participant took on an observer role that was outside the hierarchical organization. In Experiment 3, we asked participants to imagine working in an organization with two other individuals: the prosocial actor and the target of the prosocial act. Therefore, participants imagined that they were part of a single social group that would remain intact over time, which is more similar to the way in which groups actually operate in organizations. We also operationalized indirect reciprocity through helping behavior. Whereas money provides a clean measure of indirect reciprocity, it is relatively rare that a person offers cash to someone else as a reward for their generosity. However, offering time and expertise as a resource is a relatively more frequent occurrence.

Method

Experiment 3 (<https://aspredicted.org/blind.php?x=rn8sm9>) was preregistered. Any additional analyses we report are described as exploratory.

Participants. Results from an earlier study (see Supplemental Materials 2, Study 2SM⁶) that tested related predictions showed medium to large effect sizes. However, since we measure perceived selfish motive with different items in the current study, we conservatively assumed an effect size of Cohen's $d = 0.4$. Thus, we aimed to recruit 200 participants (80% power, 2-tailed). We expected some attrition due to our pre-registered exclusions, so we recruited 250 participants. The final sample for Experiment 3 was 196 Prolific workers (76.0% women; $M_{\text{age}} = 35.43$ $SD_{\text{age}} = 12.27$). We excluded individuals with duplicate IP addresses, without a valid Prolific ID, and/or who failed attention checks (they were not allowed to complete the study). These exclusions were pre-registered. Participants were paid £0.50.

Procedure. We used a 2-cell (relative power of target: higher-power vs. lower-power) between-participants design. For full experimental procedures, see Supplemental Materials 1⁵.

Participants were asked to imagine themselves as part of a Central Planning Team at a mid-sized company. They were told that there are 10 individuals on the team, including one supervisor, three project managers that report to the supervisor, and six planners, with two reporting to each project manager. An organizational chart with names in each of the roles was included below this verbal description. Participants' own role was highlighted in green as one of the three project managers (Figure 1).

Next, participants read about how the team worked and their role within it, which involved planning logistical projects for their company. As a part of this information, they read that their supervisor conducts annual evaluations that determine the project managers'

⁶ https://osf.io/h2uf6/?view_only=4f47b9604eed4f9188cdba57256a57b8

future pay and bonus. Furthermore, each of the project managers evaluates his/her two planners, and this determines the planners' future pay and bonus.

Next, they read about a situation involving two of their co-workers, Rob and Sean, and they were shown the organizational chart with Rob and Sean's roles highlighted (Figure 2). They read that Rob (another project manager) had volunteered at a homeless shelter last weekend because he had heard that Sean would be volunteering too. In the *higher-power target* condition, Sean was Rob's supervisor. In the *lower-power target* condition, Sean was one of Rob's subordinates. Therefore, the hierarchical counterpart (i.e., target) volunteered alongside the prosocial actor, and witnessed his prosocial behavior.

Measures. We asked the following dependent measures.

Perceived Motive. To measure the perceived selfless versus selfish nature of the motive, we aggregated eight items used in prior research, and which correspond to the three categories of motives suggested to drive prosocial behavior (Ariely et al., 2009; Bénabou & Tirole, 2006). These included two items probing for image-based motives and two items for extrinsic motives (both adapted from Kunstman et al., 2018). Two items measured selfless or intrinsic motive (adapted from Berman et al., 2015). Finally, two items probed general self-interest (adapted from Derfler-Rozin et al., 2018), as there is precedent for testing selfish motive in this way (Derfler-Rozin et al., 2018; Inesi et al., 2012; Lin-Healy & Small, 2012). See Supplemental Materials 1⁷ for details on the items.

A factor analysis (Direct Oblimin rotation) revealed a single factor, with all items loading at greater than .6. We therefore combined the eight items into a single scale, with higher numbers indicating more selfish motive and lower numbers indicating more selfless motives (i.e., selfless items reverse-scored; $\alpha = .93$).

⁷ https://osf.io/h2uf6/?view_only=4f47b9604eed4f9188cdba57256a57b8

Perceived Moral Character. We used an existing 12-item scale of moral character (Barasch, Levine, Berman, & Small, 2014; Berman et al., 2015) to measure Rob's perceived moral character (e.g., altruistic, immoral; 1 = not at all, 5 = extremely). Combining all twelve items revealed good internal consistency ($\alpha = .93$). An exploratory factor analysis (Direct Oblimin rotation) revealed two factors, however (all loadings $> .6$), corresponding to positive and negative moral character. We therefore created composites of the six positive moral character items ($\alpha = .87$) and the six negative moral character items ($\alpha = .92$) for the purposes of exploratory analyses.

Indirect Reciprocity. We measured indirect reciprocity by describing a scenario that happened at a later date. In this scenario, Rob needed help that the participant could provide. However, doing so would mean the participant had to stay late at work. We asked how likely and how willing the participant would be to offer Rob help (1 = not at all, 5 = extremely). The two items were significantly correlated ($r(196) = .84, p < .001; \alpha = .91$), so we combined them to form a single composite.

As manipulation checks, we asked how much power participants believed that Sean, Rob, and they themselves had in the company (1 = very little power, 5 = a great deal of power).

Results

Manipulation Checks. To test whether the power difference between the focal actors was perceived by participants as intended, we conducted a mixed ANOVA with the two targets - Rob and Sean - as repeated measures and relative power condition as a between-subject variable. There was a significant main effect of relative power condition, $F(1, 194) = 252.57, p < .001, \eta_p^2 = .57, d = 2.28$, and target, $F(1, 194) = 5.07, p = .025, \eta_p^2 = .025, d = 0.32$. The interaction between relative power condition and power was also significant, $F(1, 194) = 592.21, p < .001, \eta_p^2 = .75$. As expected, Sean ($M_{Sean} = 4.51, SD = .67$), was seen as

having more power than Rob ($M_{Rob} = 3.24$, $SD = .63$) in the higher-power target condition, $t(95) = 18.10$, $p < .001$, $d = 1.85$) while the opposite was true in the lower-power target condition ($M_{Sean} = 1.78$, $SD = .76$; $M_{Rob} = 3.31$, $SD = .78$), $t(99) = -16.92$, $p < .001$, $d = -1.69$).

We also tested whether participants saw themselves as having the same amount of power as Rob, as intended. We again ran a mixed ANOVA with role (own role vs. Rob) as a repeated measure and the relative power of the target as a between-subjects variable. None of the effects reached significance, although there was a marginal trend for participants to report both themselves and Rob as having more power in the lower-power compared to the higher-power target condition, $F(1, 194) = 3.19$, $p = .076$, $\eta_p^2 = .02$, $d = 0.26$

Perceived motive. A one-way ANOVA on perceived motive revealed a main effect of the relative power of the target, $F(1, 194) = 62.13$, $p < .001$, $\eta_p^2 = .24$, $d = 1.13$. Participants were more likely to believe that the act was selfishly rather than selflessly motivated when it was targeted towards a higher-power ($M = 4.59$, $SD = 1.23$) versus a lower-power ($M = 3.27$, $SD = 1.11$) individual.

Perceived moral character. A one-way ANOVA on the 12-item perceived moral character composite revealed a main effect of the relative power of the target, $F(1, 194) = 49.72$, $p < .001$, $\eta_p^2 = .20$, $d = 1.01$. Participants believed the prosocial actor had a lower moral character when the action was directed towards a higher- ($M = 3.28$, $SD = .66$) versus lower-power target ($M = 3.91$, $SD = .60$).

Exploratory one-way ANOVAs on the positive and negative moral character composites also revealed significant main effects, $F(1, 194) = 37.23$, $p < .001$, $\eta_p^2 = .16$, $d = 0.88$ and $F(1, 194) = 42.37$, $p < .001$, $\eta_p^2 = .18$, $d = 0.93$ respectively. Prosocial behavior that was targeted at a higher- versus lower-power individual led to perceptions of lesser positive

moral character ($M_{HP} = 2.96, SD = .66; M_{LP} = 3.53, SD = .64$) and greater negative moral character ($M_{HP} = 2.41, SD = .81; M_{LP} = 1.71, SD = .70$).

Indirect reciprocity. A one-way ANOVA on indirect reciprocity revealed a main effect of the relative power of the target, $F(1, 194) = 12.58, p < .001, \eta_p^2 = .06, d = 0.51$. Participants were less willing to help their work colleague when the prosocial actor targeted the prosocial behavior toward a higher- ($M = 3.31, SD = .94$) versus a lower-power individual ($M = 3.79, SD = .94$).

Indirect effects. We used PROCESS Model 6 (10,000 iterations; Hayes, 2017) to test our full model: that the relative power of the target of prosocial behavior affects beliefs about the actor's motives, which affects the perceived moral character of the actor, ultimately driving indirect reciprocity.

The two-stage indirect effect using the full perceived moral character scale was not significant [$-.12, SE = .08, 95\% CI = -.26, .04$]. Exploratory analyses substituting the positive moral character items for the 12-item measure revealed a significant effect [$-.16, SE = .06, 95\% CI = -.30, -.05$]. The two-stage indirect effect was not significant through the negative moral character composite [$-.01, SE = .07, 95\% CI = -.13, .15$]. See Figures 3-5 for path models, including coefficients.

Discussion. The results provide broad support for both our predicted effects as well as the underlying mechanism. Two unexpected findings emerged from Experiment 3. First, we found that moral character is best represented by two factors. Second, and relatedly, although the power relationship between the prosocial actor and their target affects both positive and negative moral character assessments, only the former drove indirect reciprocity allocations. In Supplemental Materials 2⁸, we report the results of two additional studies, which test

⁸ https://osf.io/h2uf6/?view_only=4f47b9604eed4f9188cdba57256a57b8

similar variables and process models as in Experiment 3. The results of a meta-analysis across these three experiments reveal robust support for the notion that power differences affect perceived selfish motive, which affects perceptions of moral character. However, consistent with the results presented here, perceptions of positive moral character appear to have a stronger effect on indirect reciprocity allocations than perceptions of negative moral character. This pattern makes intuitive sense in that indirect reciprocity functions as a reward for actors whose actions signal greater positive moral character (Nowak & Sigmund, 2005). Although the relative power of the target of a prosocial act vis a vis the actor did affect perceptions of negative moral character, this was simply less relevant in driving reward-based decisions (Delgado, Frank, & Phelps, 2005). In the General Discussion, we consider how the results may have differed if one or more conditions contained an act that was not prosocial, and if observers had been given the opportunity to punish.

General Discussion

There is ample evidence that the possibility of receiving indirect reciprocity increases actors' propensity to act in prosocial ways. At the same time, we know little about the *provision* of indirect reciprocity, including when and why observers reward those who cooperate. Across three experiments using different paradigms, we shed light on this dynamic. In Experiment 1, we found that people offered less indirect reciprocity to actors who helped more powerful targets compared to those who helped less powerful targets. We also included a condition in which no prosocial behavior was offered, illustrating that prosocial acts targeted up the power hierarchy received a considerable discount in indirect reciprocity compared to those directed downwards. We replicated this effect in Experiment 2 using a longitudinal paradigm in which participants observed a group's interactions across four days before receiving the key manipulation. Finally, Experiment 3 shed light on the psychological mechanisms underlying these effects. Specifically, we found that when

employees target their prosocial acts towards higher-powered others, they are seen by observers on their team as less selflessly motivated. This led to judgments of lower positive moral character and less indirect reciprocity being offered.

Theoretical Contribution

Our findings offer noteworthy contributions to a number of research areas. Most centrally, our work elucidates how observers allocate indirect reciprocity. People often witness or hear about a generous act being offered, but possess little information about the actor's past social behavior. They also lack insight in the actor's moral character, and cannot readily ascertain his or her motives. Many existing research paradigms provide observers with hidden information (e.g., speed of decision-making; Jordan et al., 2016) to understand how they use it to allocate indirect reciprocity. But what happens when such information is not known? We demonstrate that features of the social context, which are visible and widely known, can be used as signals by observers. Specifically, we show that observers use the relative power of the prosocial actor vis-à-vis his or her target to ascertain the prosocial actor's motive, assess his or her moral character, and ultimately decide how much indirect reciprocity to offer.

Second, by identifying power differences as a signal that observers use, our work has the potential to explain variations in indirect reciprocity whenever hierarchy is salient. Power differences are an inherent feature of social life (Emerson, 1962). Even when such information is not made explicit to observers, they still tend to encode power differences among individuals (Zitek & Tiedens, 2012). In this way, power may predict indirect reciprocity allocations across a variety of situations.

Our work also contributes to the understanding of power hierarchies. The vast majority of research on power focuses on one of two areas: how people *gain* positions of higher versus lower power (e.g., Cheng, Tracy, Foulsham, Kingstone, & Henrich, 2013), and

the experience of *holding* that position (e.g., Galinsky, Rucker, & Magee, 2015; Keltner, Gruenfeld, & Anderson, 2003; Whitson et al., 2013). Our work falls into a third area: How power hierarchies affect observers' understanding of a social situation. Power is readily observable (Magee & Galinsky, 2008; Berger, Cohen, & Zelditch, 1972). At the same time, we know relatively little about how it affects observers' understanding of events. Early work theorized as to its importance, especially in the context of compliments and other forms of "cheap talk" (Jones, 1964), yet empirical research remains scarce (cf. Derfler-Rozin et al., 2018; Overbeck et al., 2006). Here, we demonstrate that the perception of power hierarchies affects observers' indirect reciprocity allocations, a fundamental element of cooperation in groups.

Finally, our work provides a nuanced understanding of observers' psychology as they calculate and calibrate the indirect reciprocity that they offer prosocial actors. Although some attention has been paid to the importance of perceived intent in allocating indirect reciprocity (Hoffman et al., 2015; Jordan et al., 2016), most empirical work focuses on magnitude: more prosocial acts leads to more indirect reciprocity being awarded. Instead, our work looks at a single act, and is the first to identify and document empirically the importance of motive attributions and moral character assessments in explaining differences in indirect reciprocity allocations.

Practical Contribution

Our research offers practical implications for those interested in maximizing indirect reciprocity within their social group. Research has highlighted the importance of giving to others as a means of ultimately securing one's own success in organizations (e.g., Grant, 2013). For example, generous individuals are more likely to receive rewards by encouraging reciprocity from others, by being seen as a nice person, and by establishing stronger networks. The current set of results adds a critical piece to this puzzle: it is not only the

amount of giving that matters, but also to whom the prosocial behavior is directed that affects downstream payoffs. For those looking to receive credit for their prosocial acts, it might be useful for prosocial actors to consider how their intended beneficiary – and specifically their relative power – will influence observers' perceptions. Our work suggests that it is important to be mindful when giving to those with more power. Strategically, it may be smart to give up in private, but give down in public. More generally, managers should consider the ways in which power hierarchies might inadvertently influence both the ways in which their prosocial acts are being judged, as well as their judgments of others.

Limitations and Future Research Directions

One finding that emerged across Experiment 3 and a meta-analysis reported in Supplemental Materials 2⁹ is that indirect reciprocity allocations are more strongly driven by participants' assessments of positive moral character traits than by assessments of negative moral character traits. One reason for this effect may be that participants in our experiments were given the opportunity to reward (i.e. offer indirect reciprocity), but not to punish. Prior research has shown that reward-relevant behavior is predicted by perceptions of the actors' positive - but not negative - moral character traits (Delgado, Frank, & Phelps, 2005). Indirect reciprocity is defined as a reward for behavior that signals the presence of positive moral character (Nowak & Sigmund, 2005). Thus, participants' beliefs about the actor's positive moral character are more relevant than their beliefs about the actor's negative moral character. An interesting avenue for future research, therefore, would be to consider observers' decisions to punish and understand the extent to which negative moral character ascriptions drive these decisions. Existing research shows that negative moral character ascriptions beget harsh judgments from observers (Pizarro & Tannenbaum, 2012) and that

⁹ https://osf.io/h2uf6/?view_only=4f47b9604eed4f9188cdba57256a57b8

observers respond to selfish behavior with altruistic punishment (Fehr & Fischbacher, 2004). It seems plausible, then, that observers would be more likely to base their punishment decisions on assessments of the actors' negative moral character than on assessments of positive moral character.

One limitation of the current set of findings is that they were all conducted using online samples. This potentially limits the generalizability of the results to other populations. Related to this, we led participants to believe that the prosocial actor and the beneficiaries were real people, when in fact they were not. An important avenue for future research, therefore, would be to demonstrate that similar patterns emerge in laboratory and field settings, and when all actors are real people.

Future research could also consider the implications of these patterns for prosocial actors. Prior research has demonstrated that prosocial actors are more likely to cooperate when their actions will be known to observers because they are more likely to receive indirect reciprocity (e.g., Ariely et al., 2009; Jordan et al., 2016; Milinski et al., 2002). The results of the present research show that *observers* are more likely to reward downward-directed acts of prosociality over upward-directed acts. Therefore, it is possible that prosocial actors will be more likely to give to lower- versus higher-power beneficiaries when their actions are public versus private. Relatedly, do actors accurately understand that the relative power of their intended beneficiary will influence how they are perceived? Although these questions are beyond the scope of the current research, they present fascinating and important avenues for future research.

In Experiment 3, power differences between prosocial actors and their targets affected indirect reciprocity by altering assessments of the prosocial actors' positive moral character (informed by attributions of motive). The notion that indirect reciprocity allocations are linked to moral character assessments is consistent with theories of how and why indirect

reciprocity exists: groups benefit from rewarding and supporting members who have high positive moral character, so we are motivated to offer them greater benefits (Nowak & Sigmund, 2005; Wedekind & Milinski, 2000). At the same time, it is possible that more calculative reasons underlie observers' tendency to offer less resources to those who target generosity at higher- versus lower-power individuals. That is, higher-power targets have more valuable resources at their disposal than lower-power targets. Since prosocial actors can expect direct reciprocity from their targets, then actors who target higher-power individuals can expect to earn more valuable direct reciprocity in comparison to those who target lower-power individuals. Observers' indirect reciprocity allocation decisions may reflect a desire to "balance accounts", meaning that they offer less to actors who targets higher-power beneficiaries. Exploring this alternative mechanism is an interesting avenue for future research.

Additionally, research could examine the robustness of this effect when considering different contextual factors. For example, the effect observed across our studies might diminish or even disappear if the actor brags about it (Berman et al., 2015). Since bragging signals a selfish motive to gain social credit, then observers would attribute the act – no matter whom the target – to less altruistic motives, driving their indirect reciprocity decisions. Also, the effect of power differences on indirect reciprocity might be weaker if the actor performs a kind act, but does not intend for the target to know about it. A study that we report in our supplemental materials (Study 2SM in Supplemental Materials 2¹⁰) offers suggestive evidence. The paradigm was similar to Experiment 3, except we added an additional condition in which the prosocial actor did not intend for target to know about their generous

¹⁰ https://osf.io/h2uf6/?view_only=4f47b9604eed4f9188cdba57256a57b8

act. In this study, we observed a significant interaction between relative power of the target and intent condition.

Another feature that remained constant across all of our experiments was the stated gender of the actor: he was always described as male. We encourage future research to vary the gender of the involved parties, since it is possible that men and women would receive different amounts of indirect reciprocity for performing the same prosocial act. On one hand, women may receive less indirect reciprocity than men when they offer kindness to higher-power targets because such “ambitious” behavior goes against the prescription for women to be warm (Heilman, 2012; Rudman & Glick, 1999). At the same time, shifting standards (Biernat & Vescio, 2002) suggest that male prosocial actors may reap more indirect reciprocity than female prosocial actors when they give to lower-power targets, because they will be judged as kinder. Ultimately, this would yield higher levels of indirect reciprocity for men for performing the same act.

One final area for future research would be to investigate how prosocial acts directed laterally in the power hierarchy are rewarded with indirect reciprocity compared to upward- and downward-directed acts. Across our three studies, we compared upward-directed prosocial acts to downward-directed ones. Our theory assumes that upward-directed acts are discounted because they are seen to be driven by less altruistic motives. Since lateral relationships are generally characterized by lower levels of dependence, we predict that observers would be less likely to discount lateral prosocial acts compared to upward-directed acts, and would allocate relatively equal amounts compared to downward-directed acts. Future research could investigate this possibility, in addition to exploring and or manipulating the degree of interdependence between lateral actors in a power hierarchy.

Conclusion

Indirect reciprocity is a critical process underlying cooperation in social groups. However, we know little about the dynamics that determine when and why it flows from one individual to another. Across three experiments, we find that power differences between prosocial actors and their targets influence the amount of indirect reciprocity that observers offer. In doing so, we show not only that contextual social signals are important in indirect reciprocity, but also that power, a central feature of social life, acts as one of these signals.

References

- Alexander, R. D. (1987). *The biology of moral systems*. New York, NY: Aldine De Gruyter.
- Ariely, D., Bracha, A., & Meier, S. (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *American Economic Review*, 99(1), 544-55. <http://dx.doi.org/10.1257/aer.99.1.544>
- Baker, W. E., & Bulkley, N. (2014). Paying it forward vs. rewarding reputation: Mechanisms of generalized reciprocity. *Organization Science*, 25(5), 1493-1510. <https://doi.org/10.1287/orsc.2014.0920>
- Barasch, A., Levine, E. E., Berman, J. Z., & Small, D. A. (2014). Selfish or selfless? On the signal value of emotion in altruistic behavior. *Journal of Personality and Social Psychology*, 107(3), 393-413. <http://dx.doi.org/10.1037/a0037207>
- Batson, C. D. (1998). Altruism and prosocial behavior. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, pp. 282-316). Boston: McGraw-Hill.
- Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5), 1652-1678. <http://dx.doi.org/10.1257/aer.96.5.1652>
- Berger, J., Cohen, B. P., & Zelditch, M. (1972). Status characteristics and social interaction. *American Sociological Review*, 37(3), 241-255. <http://dx.doi.org/10.2307/2093465>
- Berman, J. Z., Levine, E. E., Barasch, A., & Small, D. A. (2015). The braggart's dilemma: On the social rewards and penalties of advertising prosocial behavior. *Journal of Marketing Research*, 52(1), 90-104. <http://dx.doi.org/10.1509/jmr.14.0002>
- Biernat, M., & Vescio, T.K. (2002). She swings, she hits, she's great, she's benched: Implications of gender-based shifting standards for judgment and behavior. *Personality and Social Psychology Bulletin*, 28(1), 66-77. <https://dx.doi.org/10.1177/0146167202281006>

- Bolton, G. E., Brandts, J., & Ockenfels, A. (1998). Measuring motivations for the reciprocal responses observed in a simple dilemma game. *Experimental Economics*, *1*(3), 207-219. <http://dx.doi.org/10.1023/A:1009951108693>
- Brooks, A. W., & Schweitzer, M. E. (2011). Can Nervous Nelly negotiate? How anxiety causes negotiators to make low first offers, exit early, and earn less profit. *Organizational Behavior and Human Decision Processes*, *115*(1), 43-54. <http://dx.doi.org/10.1016/j.obhdp.2011.01.008>
- Carlson, R. W., & Zaki, J. (2018). Good deeds gone bad: Lay theories of altruism and selfishness. *Journal of Experimental Social Psychology*, *75*, 36-40. <https://dx.doi.org/10.1016/j.jesp.2017.11.005>
- Chartrand, T. L., van Baaren, R. B., & Bargh, J. A. (2006). Linking automatic evaluation to mood and information processing style: Consequences for experienced affect, impression formation, and stereotyping. *Journal of Experimental Psychology: General*, *135*(1), 70-77. <http://dx.doi.org/10.1037/0096-3445.135.1.70>
- Cheng, J. T., Tracy, J. L., Foulsham, T., Kingstone, A., & Henrich, J. (2013). Two ways to the top: Evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. *Journal of Personality and Social Psychology*, *104*(1), 103-125. <http://dx.doi.org/10.1037/a0030398>
- Critcher, C. R., & Dunning, D. (2011). No good deed goes unquestioned: Cynical reconstruals maintain belief in the power of self-interest. *Journal of Experimental Social Psychology*, *47*(6), 1207-1213. <https://doi.org/10.1016/j.jesp.2011.05.001>
- Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin*, *90*(2), 272-292. <https://dx.doi.org/10.1037/0033-2909.90.2.272>

- Deaux, K., & LaFrance, M. (1998). Gender. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 788–827). Boston: McGraw-Hill.
- Delgado, M., Frank, R. & Phelps, E. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, 8, 1611–1618. <https://doi.org/10.1038/nn1575>
- Derfler-Rozin, R., Baker, B., & Gino, F. (2018). Compromised ethics in hiring processes? How referrers' power affects employees' reactions to referral practices. *Academy of Management Journal*, 61(2), 615-636. <http://dx.doi.org/10.5465/amj.2015.1355>
- Emerson, R. M. (1962). Power-dependence relationships. *American Sociological Review*, 27(1), 31– 41. <http://dx.doi.org/10.2307/2089716>
- Evans, A. M., & Van de Calseyde, P. P. (2017). The effects of observed decision time on expectations of extremity and cooperation. *Journal of Experimental Social Psychology*, 68, 50-59. <http://dx.doi.org/10.1016/j.jesp.2016.05.009>
- Fast, N. J., Gruenfeld, D. H., Sivanathan, N., & Galinsky, A. D. (2009). Illusory control: A generative force behind power's far-reaching effects. *Psychological Science*, 20(4), 502-508. <http://dx.doi.org/10.1111/j.1467-9280.2009.02311.x>
- Fehr, E. & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, 25 (2), 63-87. [https://doi.org/10.1016/S1090-5138\(04\)00005-4](https://doi.org/10.1016/S1090-5138(04)00005-4).
- Fein, S. (1996). Effects of suspicion on attributional thinking and the correspondence bias. *Journal of Personality and Social Psychology*, 70(6), 1161-1184. <http://dx.doi.org/10.1037/0022-3514.70.6.1164>
- Funder, D. C. (2004). *The personality puzzle* (3rd ed.). New York, NY: W. W. Norton.
- Galinsky, A. D., Rucker, D. D., & Magee, J. C. (2015). *Power: Past findings, present considerations, and future directions. APA handbook of personality and social*

psychology: Vol. 1. Attitudes and social cognition. Washington, DC: American Psychological Association.

Galinsky, A. D., Rus, D., & Lammers, J. (2011). Power: A central force governing psychological, social, and organizational life. In D. De Cremer, R. van Dick, & J. K. Murnighan (Eds.), *Social psychology and organizations* (pp. 17–38). New York, NY: Routledge.

Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189–211). New York, NY: Guilford Press.

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, 25(2), 161–178. <http://dx.doi.org/10.2307/2092623>

Grant, A. M. (2013). *Give and take: A revolutionary approach to success*. New York, NY: Viking Press.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach (2nd ed.)*. New York, NY: Guilford Publications.

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior*, 32, 113–135. <https://dx.doi.org/10.1016/j.riob.2012.11.003>

Herne, K., Lappalainen, O., & Kestila-Kekkonen, E. (2013). Experimental comparison of direct, general, and indirect reciprocity. *The Journal of Socio-Economics*, 45, 38–46. <http://dx.doi.org/10.1016/j.socec.2013.04.003>

Hoffman, M., Yoeli, E., & Nowak, M. A. (2015). Cooperate without looking: Why we care what people think and not just what they do. *Proceedings of the National Academy of Sciences*, 112(6), 1727–1732. <https://dx.doi.org/10.1073/pnas.1417904112>

- Inesi, M. E., Gruenfeld, D. H., & Galinsky, A. D. (2012). How power corrupts relationships: Cynical attributions for others' generous acts. *Journal of Experimental Social Psychology, 48*(4), 795-803. <http://dx.doi.org/10.1016/j.jesp.2012.01.008>
- Jones, E. E. (1964). *Ingratiation: A social psychological analysis*. New York, NY: Appleton-Century-Crofts.
- Jordan, J. J., Hoffman, M., Nowak, M. A., & Rand, D. G. (2016). Uncalculating cooperation is used to signal trustworthiness. *Proceedings of the National Academy of Sciences, 113*(31), 8658-8663. <http://dx.doi.org/10.1073/pnas.1601280113>
- Kelley, H. H. (1972). Causal schemata and the attribution process. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 151-174). Morristown, NJ: General Learning Press.
- Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review, 110*(2), 265-284. <http://dx.doi.org/10.1037/0033-295X.110.2.265>
- Kunstman, J. W., Fitzpatrick, C. B., & Smith, P. K. (2018). Poisoned praise: Discounted praise backfires and undermines subordinate impressions in the minds of the powerful. *Social Psychological and Personality Science, 9*(4), 470-480. <http://dx.doi.org/10.1177%2F1948550617712028>
- Leimar, O., & Hammerstein, P. (2001). Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 268*, 745-753. <http://dx.doi.org/10.1098/rspb.2000.1573>
- Lin-Healy, F., & Small, D. A. (2012). Cheapened altruism: Discounting personally affected prosocial actors. *Organizational Behavior and Human Decision Processes, 117*(2), 269-274. <http://dx.doi.org/10.1016/j.obhdp.2011.11.006>

- Magee, J. C., & Galinsky, A. D. (2008). Social hierarchy: The self-reinforcing nature of power and status. *Academy of Management Annals*, 2(1), 351-398.
<http://dx.doi.org/10.5465/19416520802211628>
- McNeely, B. L., & Meglino, B. M. (1994). The role of dispositional and situational antecedents in prosocial organizational behavior: An examination of the intended beneficiaries of prosocial behavior. *Journal of Applied Psychology*, 79(6), 836-844.
<http://dx.doi.org/10.1037/0021-9010.79.6.836>
- Milinski, M., Semmann, D., & Krambeck, H. J. (2002). Reputation helps solve the 'tragedy of the commons'. *Nature*, 415, 424-426. <http://dx.doi.org/10.1038/415424a>
- Miller, D. T. (1999). The norm of self-interest. *American Psychologist*, 54(12), 1053-1060.
<https://doi.org/10.1037/0003-066X.54.12.1053>
- Newman, G. E., & Cain, D. M. (2014). Tainted altruism: When doing some good is evaluated as worse than doing no good at all. *Psychological Science*, 25(3), 648-655.
<http://dx.doi.org/10.1177/0956797613504785>
- Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature*, 393, 573-577. <http://dx.doi.org/10.1038/31225>
- Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437, 1291-1298. <http://dx.doi.org/10.1038/nature04131>
- Overbeck, J. R., Tiedens, L. Z., & Brion, S. (2006). The powerful want to, the powerless have to: Perceived constraint moderates causal attributions. *European Journal of Social Psychology*, 36(4), 479-496. <http://dx.doi.org/10.1002/ejsp.353>
- Pettigrew, T.F. (1979). The ultimate attribution error: Extending Allport's cognitive analysis of prejudice. *Personality and Social Psychology Bulletin*, 5(4), 461-476.
<http://dx.doi.org/10.1177/014616727900500407>

- Pizarro, D. A., & Tannenbaum, D. (2012). Bringing character back: How the motivation to evaluate character influences judgments of moral blame.
- Reeder, G. D., Kumar, S., Hesson-McInnis, M. S., & Trafimow, D. (2002). Inferences about the morality of an aggressor: The role of perceived motive. *Journal of Personality and Social Psychology*, 83(4), 789–803. <http://dx.doi.org/10.1037/0022-3514.83.4.789>
- Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: the hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology*, 77(5), 1004-1010. <http://dx.doi.org/10.1037/0022-3514.77.5.1004>
- Sekaquaptewa, D., & Espinoza, P. (2004). Biased processing of stereotype-incongruity is greater for low than high status groups. *Journal of Experimental Social Psychology*, 40(1), 128-135. [https://dx.doi.org/10.1016/S0022-1031\(03\)00093-3](https://dx.doi.org/10.1016/S0022-1031(03)00093-3)
- Simpson, B., & Willer, R. (2008). Altruism and indirect reciprocity: The interaction of person and situation in prosocial behavior. *Social Psychology Quarterly*, 71(1), 37-52. <http://dx.doi.org/10.1177/019027250807100106>
- Van Vugt, M., Roberts, G., & Hardy, C. (2007). Competitive altruism: Development of reputation-based cooperation in groups. In R. Dunbar & L. Barrett (Eds.), *Handbook of evolutionary psychology* (pp. 531-540). Oxford, England: Oxford University Press.
- Wedekind, C. & Braithwaite, V. A. (2002). The long-term benefits of human generosity in indirect reciprocity. *Current Biology*, 12(12), 1012-1015. [http://dx.doi.org/10.1016/S0960-9822\(02\)00890-4](http://dx.doi.org/10.1016/S0960-9822(02)00890-4)
- Wedekind, C., & Milinski, M. (2000). Cooperation through image scoring in humans. *Science*, 288, 850-852. <http://dx.doi.org/10.1126/science.288.5467.850>
- Whitson, J. A., Liljenquist, K. A., Galinsky, A. D., Magee, J. C., Gruenfeld, D. H., & Cadena, B. (2013). The blind leading: Power reduces awareness of constraints.

Journal of Experimental Social Psychology, 49(3), 579-582.

<http://dx.doi.org/10.1016/j.jesp.2012.10.009>

Zahavi, A., & Zahavi, A. (1997). *The handicap principle: A missing piece of Darwin's puzzle*.

New York, NY: Oxford University Press.

Zitek, E. M., & Tiedens, L. Z. (2012). The fluency of social hierarchy: the ease with which

hierarchical relationships are seen, remembered, learned, and liked. *Journal of*

Personality and Social Psychology, 102(1), 98-115.

<http://dx.doi.org/10.1037/a0025345>