COGNITIVE SCIENCE A Multidisciplinary Journal



Cognitive Science 43 (2019) e12773 © 2019 Cognitive Science Society, Inc. ISSN: 1551-6709 online DOI: 10.1111/cogs.12773

When and Why People Evaluate Negative Reciprocity as More Fair Than Positive Reciprocity

Alex Shaw, Anam Barakzai, Boaz Keysar

Department of Psychology, University of Chicago

Received 27 May 2018; received in revised form 6 June 2019; accepted 13 June 2019

Abstract

If you are kind to me, I am likely to reciprocate and doing so feels fair. Many theories of social exchange assume that such reciprocity and fairness are well aligned with one another. We argue that this correspondence between reciprocity and fairness is restricted to interpersonal dyads and does not govern more complex multilateral interactions. When multiple people are involved, reciprocity leads to partiality, which may be seen as unfair by outsiders. We report seven studies, conducted with people from the United States, in which participants were asked to evaluate situations involving resource distribution in contexts such as economic games, government, and the workplace. Specifically, we find that equal resource distribution in multilateral interactions is seen as more fair than engaging in reciprocity. We also find that negative reciprocity is seen as more fair than positive reciprocity in these multilateral situations and demonstrate that there are contexts where favoritism is not viewed as unfair. These findings are important for theories of fairness and reciprocity as they demonstrate the central role of perceived partiality in the evaluation of multi-party resource allocation.

Keywords: Fairness; Inequity Aversion; Reciprocity; Cooperation; Impartiality

1. Introduction

Reciprocity is ubiquitous in human society. For example, if Sally gives her co-worker Bob some tickets to see a band he really wants to see, he might reciprocate by helping her with some problem at work. This is classic reciprocity—Sally scratched Bob's back and Bob returns the favor. Many would also find Bob's actions to be perfectly fair; by repaying Sally's generosity, Bob has observed the norm of reciprocity and that is

Correspondence should be sent to Alex Shaw, Department of Psychology, University of Chicago, 5848 University Ave., Chicago, IL 60637, USA. E-mail: alex.w.shaw@gmail.com

perceived as fair (e.g., Baumard, André, & Sperber, 2013; Falk, Fehr, & Fischbacher, 2008; Haidt & Joseph, 2008; Pinker, 2008).

As evident as this relationship between fairness and reciprocity may seem to be, we propose that it is true for dyadic situations, but not in multilateral situations that involve multiple people and decisions to distribute group resources among multiple others (Shaw, 2013). For example, imagine that Sally was the office manager and she assigned Bob a particularly fancy vacant office, and that Bob later reciprocated by assigning Sally a luxury company car rather than assigning it to another equally deserving employee. Such favoritism is a commonplace occurrence in many offices. Indeed, according to a Georgetown University's McDonough School of Business survey, a quarter of senior executives said they use favoritism with employee promotions and 92% have witnessed such practices (Reinsch & Gardner, 2014). While this may permeate many aspects of our society, such reciprocity does not seem fair.

In seven studies, we examine people's fairness evaluations of multilateral interactions in economic games, government, and office contexts. We make two predictions based on the idea that fairness concerns are rooted in avoiding partiality (Choshen-Hillel, Shaw, & Caruso, 2015; Tyler, 2000). First, we predict that fairness and reciprocity will conflict in multilateral interactions because reciprocity will inevitably lead to unequal distribution that may be viewed as unfair, particularly if it is seen as showing favoritism. Second, we predict that people will view negative reciprocity as more fair than positive reciprocity because the former will be less likely to be seen as a form of partiality. To set up these predictions, we briefly review research on reciprocity and fairness.

Most investigations of fairness and reciprocity have examined dyadic interactions where there is no conflict between fairness and reciprocity; indeed, reciprocity often supports fairness in these contexts. A large literature has demonstrated that reciprocity is common in everyday situations and in economic games like the prisoner's dilemma and public goods games (Andreoni, 1995; Binmore, 2006; Delton, Krasnow, Cosmides, & Tooby, 2011; Falk & Fischbacher, 2006; Fehr & Gächter, 2000; Gurven, 2003, 2006; Gurven & Winking, 2008; Hill & Kaplan, 1993; Kiyonari & Barclay, 2008). Many researchers have suggested that fairness concerns support such reciprocal sharing and cooperation (Baumard et al., 2013; Chiang, 2010; Falk et al., 2008; Ma, Tunney, & Ferguson, 2014; Pinker, 2008). Positive reciprocity involves paying back someone's generosity, which promotes cooperation, while negative reciprocity results in punishing someone's selfishness, which results in suppressing selfishness. Therefore, in dyadic interactions, both positive and negative reciprocity will often lead to reducing inequalities and will frequently be seen as fair.

However, this correspondence between fairness and reciprocity may not hold in multilateral interactions because reciprocity in such interactions will often lead to more inequality rather than less inequality. In multilateral interactions, decisions cannot be easily boiled down to cooperation or defection because it will often be the case that cooperating with one person may mean defecting against another (e.g., DeScioli & Kurzban, 2013; Shaw, DeScioli, Barakzai, & Kurzban, 2017). Thus, preferential treatment based on reciprocity will often not lead to a reduction in inequality and can easily exacerbate current inequalities. Indeed, if two agents selectively cooperate with each other at the exclusion of others, this will be perfectly aligned with reciprocity but will lead to increased inequality (DeScioli & Kurzban, 2013; Murnighan, 1978; Ray, 2007; Snyder, 1984; Van Beest, Van Dijk, De Dreu, & Wilke, 2005).

Thus, reciprocity in multilateral interactions often will result in others receiving unequal outcomes. However, it is unclear from this fact alone whether reciprocity in multilateral situations will be seen as unfair, and whether positive or negative reciprocity will be viewed as equally fair or unfair.

In contexts like economic games, government, and workplaces, people often think inequality is unfair, but they do not find all inequality to be unfair (Adams, 1965; Elster, 2006). There is considerable research demonstrating that people respond negatively when they or others receive unequal outcomes, and people will sometimes take personal costs to reduce these inequalities (Damon, 1977; Dawes, Fowler, Johnson, McElreath, & Smirnov, 2007; Fehr & Schmidt, 1999; Messick, 1995; Shaw & Knobe, 2013). However, not all inequality is treated as equally unfair. People tend to regard inequality as more unfair when it is based on favoritism or partiality (Choshen-Hillel, Shaw, & Caruso, 2018; Elster, 2006; Gordon-Hecker, Rosensaft, Pittarello, Shalvi, & Bereby-Meyer, 2017; Shaw, 2013; Tyler, 2000). Unequal resource distributions are perceived as partial when they are not motivated by a socially agreed-upon justification such as merit, need, or an impartial randomization procedure, such as a lottery (Elster, 2006; Hook & Cook, 1979; Shaw & Olson, 2012, 2014; Thibaut & Walker, 1975; Tyler, 2000). For example, if one person works harder or does a better job than someone else, giving more to this person will result in inequality, but it is unlikely to be perceived as unfair or showing partiality because merit is a culturally agreed-upon justification for inequality. Therefore, inequality is not intrinsically unfair; people create inequality, and think it is fair to do so, if it is based on impartial justifications (Deutsch, 1975; Rasinski, 1987; Shaw, 2013).

The following experiments explore two questions that follow from these notions of reciprocity, equality, and impartiality: Is reciprocity seen as an acceptable justification for inequality in multilateral interactions? And will positive and negative reciprocity be seen as equally fair or unfair, that is, equally good or bad justifications for such inequality? These experiments primarily examine how people evaluate others' decisions.

Our first prediction is that people will evaluate equal allocations, all else being equal, as more fair than reciprocity-based allocations in multilateral interactions. People frequently find equal outcomes to be impartial because such outcomes do not convey that the allocator favors one recipient over another (Choshen-Hillel et al., 2015; Kleiman-Weiner, Shaw, & Tenenbaum, 2017). Therefore, in these multilateral contexts where fairness concerns are prominent and partiality is viewed negatively, we would expect that people find equal distribution to be more fair than reciprocity, whether positive or negative, because reciprocity in these settings demands treating others unequally and in a way that is likely to seem partial.

Of course, the more interesting question is whether positive or negative reciprocity will be seen as more unfair. This leads to our second prediction: People will think that negative reciprocity is more fair than positive reciprocity because negative reciprocity is less likely to be seen as an attempt to show favoritism toward the person who receives better treatment. Negative reciprocity, or retaliation, may function as a better justification for inequality than positive reciprocity. Specifically, if people's fairness evaluations are driven by concerns with partiality, then they may evaluate the exact same inequality differently depending on whether it is based on positive or negative reciprocity. For example, imagine that Bob distributes resources unequally between Sally and Tim, and Sally ends up with the larger share. In one case, Bob gives more to Sally because Sally treated him well in the past (positive reciprocity towards Sally) and in the other case because Tim treated him poorly in the past (negative reciprocity towards Tim). When Bob engages in positive reciprocity by giving more to Sally to reciprocate for the way she treated him in the past, this inequality is likely to be seen as a form of favoritism toward Sally. However, when Bob gives more to Sally based on negative reciprocity, it is unlikely that participants will infer the same degree of favoritism toward Sally. Instead, participants will assume that Sally receiving more is merely a side effect of Bob getting even with Tim. Therefore, in multilateral contexts where impartiality is important, people may view negative reciprocity to be more fair than positive reciprocity.

In order to investigate these predictions, most of our experiments focus on contexts in which people are dividing third-party resources in multilateral interactions; in contexts like economics games, government, and workplaces where fairness concerns seem to be particularly important (Adams, 1965; Messick, 1995; Tyler, 2000). We focus on these interactions because such interactions highlight this tension between reciprocity and fairness and are commonplace (e.g., Abbink, Irlenbusch, & Renner, 2002; Barr & Serra, 2009; Brick, Palmon, & Wald, 2006; Dungan, Waytz, & Young, 2014; Harstad & Svensson, 2011; Niemi, Wasserman, & Young, 2018; Van Beest, Wilke, & Van Dijk, 2003). For example, when a government official makes a decision about which projects to fund or a boss decides which employees should get a raise, they are engaged in a multilateral interaction. In most of these cases, the decision-makers are not making a decision about whether to be selfish or generous with their own resources, but they are instead making a decision about how to distribute group resources and whether to treat others equally or to show favoritism toward particular others (Shaw & Knobe, 2013). Given how widespread these situations are and the different concerns that may underlie people's evaluations of such decision-making, it seems important to understand how people evaluate such allocations and the mechanisms that underlie their reactions.

Thus, in seven studies we examine these two predictions by exploring how people evaluate reciprocity in multilateral interactions involving distribution of third-party resources, involving scenarios about economic games (Study 1a), government (Study 2 and 3), and an office workplace (Study 4, 5, and 6). We also investigate these phenomena in a consequential economic game in our lab, in which people's resource decisions affected others' real outcomes (Study 1b). Furthermore, we examine if our proposed mechanism, perceptions of favoritism, mediates these fairness evaluations (Study 3, 4, 5, and 6). We also examine two key boundary conditions for this effect (Supplemental Study S1 and Study 5) and examine how such reciprocity influences downstream aspects such as self-reported workplace satisfaction (Study 4) and self-reported behavioral

outcomes (Study 2 and 3). Moreover, we compare positive and negative reciprocity to a well-supported justification for unequal treatment, namely merit (Study 4). Finally, we note that not all partiality should be seen as unfair (e.g., giving more to one's own child as compared to other children is likely seen as fair). Indeed, partiality may be acceptable in interpersonal scenarios when sharing one's own resources with friends (Study 5) and the contexts in which impartiality is considered to be important will likely vary by culture (we return to this issue in the General Discussion).

1.1. General methodological notes

To avoid repetition, we explain elements of the method that are common to most studies. Except for Study 1b, participants were recruited online using the Amazon Mechanical Turk (MTurk) and Turk Prime (Litman, Robinson, & Abberbock, 2016) websites. Participation was restricted to residents of the United States with an approval rating of 95% and above and participants were paid 25 cents for completing the survey.

In all studies, we decided ahead of data collection not to exclude any participants in order to avoid selection issues (we included comprehension checks in each study and the rates of failing were generally low, below ~10%). For all studies (except Study 1a), we recruited ~50 participants per cell. Further we only analyzed the results of each study after data collection was complete. Although we included only ~40–50 participants per cell in each study, we conceptually and directly replicate the two main predictions of the paper in our subsequent studies. Table 1 provides numerous replications of our main effect, which serves as "post-registration" data (Scholl, 2017). All information needed to replicate these studies is included in the manuscript. We did not collect demographic data beyond sex and age from our MTurk or in-laboratory participants, as we did not have a priori hypotheses about differences associated with participant demography. Thus, we cannot speak to the demographic diversity of our sample in terms of factors outside sex and age.

2. Study 1a

In Study 1a, we examined how people react to different types of resource distribution in a hypothetical third-party economic game. Participants read that four people participated in an economic game. In round 1 of the game, Jake allocated some resources between two recipients, giving one of them more than the other. In the positive condition, Jake gave Kate more than the other recipient, and in the negative condition he gave Kate less than the other recipient. Participants then read that, in round 2, Kate was asked to divide resources between Jake and another recipient. So, while Jake allocated between Kate and Person A in round 1, Kate allocated between Jake and Person B in round 2. We varied how Kate divided the resources. She either divided them equally between Jake and the other recipient (equality condition) or reciprocated Jake's action from round 1 (reciprocity conditions). That is, Kate either allocated equally between Jake and Person B, repaid Jake's favoritism by giving more to him in the positive reciprocity condition, or retaliated by giving him less in the negative reciprocity condition. Participants were then asked to provide a fairness evaluation and to state what they would do in Kate's place.

If fairness concerns are tracking partiality in these situations, then two predictions follow. First, equality should be seen as more fair than any reciprocity because equal division will not be seen as partial. Second, negative reciprocity should be seen as more fair than positive reciprocity. We make this prediction because the retaliation displayed by negative reciprocity is less likely to be seen as being based on partiality than the preferential treatment displayed by positive reciprocity.

2.1. Method

2.1.1. Participants

One hundred and sixty-five adults (49% females, M = 29.12 years, SD = 8.69) participated in this survey.

2.1.2. Design and procedure

Participants were randomly assigned in a 2 (valence: positive, negative) \times 2 (response: reciprocity, equality) between-participants design. Participants were told about two rounds of an economic game in which people were asked to distribute resources to others. All participants first read:

Four people (Jake, Kate, Laura, and David) were brought into the lab and were asked to make decisions about how resources should be divided. Each person could divide the resources however they wanted among the two recipients.

In the positive conditions, participants read the following [within each set of brackets the reciprocity condition appears first and the equality condition appears second]:

In the first round, Jake was given \$6 to divide among Kate and Laura. He gave \$5 to Kate and \$1 to Laura.

In the second round, Kate was given \$6 to divide among Jake and David. She gave [\$5 to Jake and \$1 to David/ \$3 to Jake and \$3 to David]. There are only two rounds and all four people know what happened in round 1.

Participants in the negative conditions read the following [within each set of brackets the reciprocity condition appears first and the equality condition appears second]:

In the first round, Jake was given 6 to divide among Kate and Laura. He gave 1 to Kate and 5 to Laura.

6 of 35

In the second round, Kate was given \$6 to divide among Jake and David. She gave [\$1 to Jake and \$5 to David/ \$3 to Jake and \$3 to David]. There are only two rounds and all four people know what happened in round 1.

Note, in all conditions, we explicitly said that "all four people know what happened in round 1" to make it clear that everyone had common knowledge of others' previous choices. Participants then answered two fairness questions and a choice question about how they would behave in Kate's situation. In the fairness questions, participants rated their agreement with two statements about Kate's actions, "I agree with what Kate did" and "What Kate did was fair," on a 7-point Likert scale that went from 1 (*Strongly disagree*) to 7 (*Strongly agree*). These items were presented in a random order and were highly correlated, r(165) = .83, p < .001, so we combined them by taking the mean of the two measures and report a "general fairness" measure (all combined measures in subsequent studies in the manuscript are a mean of the measur Do children think the tyranny of the majority is unfairness).

Participants then answered the choice question which asked what they themselves would have done in Kate's place in the present scenario. In the positive conditions, participants were told, "Remember, Jake was the one who gave Kate \$5 and Laura \$1 in round 1," and in the negative conditions they were told, "Remember, Jake was the one who gave Kate \$1 and Laura \$5 in round 1." In both conditions, participants were asked, "If you were Kate, how much would you want to give to Jake and David in round 2?" The options were to "give \$3 to Jake and \$3 to David," "\$5 to Jake and \$1 to David," and "\$1 to Jake and \$5 to David."

2.2. Results

2.2.1. Fairness

A 2 (valence: positive, negative) × 2 (response: reciprocity, equality) ANOVA on the fairness ratings revealed a main effect of valence, F(1, 161) = 10.17, p = .002, $\eta_p^2 = .059$. This result indicates that participants thought that Kate was more fair in the negative conditions (M = 5.60, SD = 1.38) than the positive conditions (M = 5.00, SD = 1.78). We also found a main effect of response, F(1, 161) = 45.6, p < .001, $\eta_p^2 = .221$, such that participants in the equality conditions saw Kate's actions as more fair (M = 6.02, SD = 1.12) than participants in the reciprocity conditions (M = 4.60, SD = 1.72). Valence and response also interacted significantly, F(1, 161) = 7.98, p = .005, $\eta_p^2 = .047$, which we followed up on with planned comparisons. We compared fairness ratings for the different valence levels first in the reciprocity,

We compared fairness ratings for the different valence levels first in the reciprocity, then equality conditions. As predicted, negative reciprocity (M = 5.22, SD = 1.55) was perceived as more fair than positive reciprocity (M = 3.91, SD = 1.65), t(82) = 3.73, p < .001, d = .82. That is, participants thought the same inequality was more fair when it resulted from negative rather than positive reciprocity. Further, participants in the equality conditions rated Kate's actions as similarly fair in the negative (M = 6.06, SD = .98) and

positive condition (M = 5.99, SD = 1.24), t(79) = .31, p = .760, d = .61. Fig. 1 shows the interaction results from differential fairness ratings only in the reciprocity conditions.

2.2.2. Choice

We conducted a planned contrast Fisher's exact test to compare people's tendency to choose reciprocity vs. equal allocation in the positive and negative conditions. For these analyses, we ignore anti-reciprocity (selecting an action that was the opposite of what reciprocity would dictate, that is —giving more to Jake in round 2 when he had given Kate less in round 1), However, we include this information in the goodness of fit tests reported below. When imagining being Kate, participants were more likely to report they would reciprocate in the negative conditions (58% reciprocity vs. 42% equal) than in the positive conditions (19% reciprocity vs. 81% equal), p < .001.

We further conducted chi-squared goodness of fit tests on people's choices. In the negative conditions, participants picked reciprocity (45 out of 81, 55%) more often than equal division (33 out of 81, 41%) or anti-reciprocity (3 out of 81, 4%), $\chi^2(2, N = 81) = 34.67$, p < .001. In contrast, in the positive conditions, they picked equal giving (61 out of 84, 72.5%) more often than reciprocity (17 out of 84, 16.5%) or anti-reciprocity (11 out of 84, 12%), $\chi^2(2, N = 84) = 55.15$, p < .001.

The judgment and choice data converge and suggest that, in multilateral interactions, people view retaliation via negative reciprocity as significantly more fair than favoritism via positive reciprocity. Additionally, people evaluated equal sharing to be the most fair way of dividing resources in this context.

3. Study 1b

Study 1a found that participants thought that it was more fair to engage in negative rather than positive reciprocity. However, given that allocations in Study 1a did not affect



Fig. 1. Mean fairness rating by condition in Study 1a (hypothetical economic game) and 2 (consequential economic game). Bars represent SE's of the mean.

9 of 35

the subject, it may be argued that people were not treating the distributions as meaningful. In Study 1b, we examined how people respond to decisions that affect them. Study 1b was similar to Study 1a except that participants actually played a role similar to Kate's (the second decision-maker) in Study 1a, and their decisions involved real monetary consequences. We expected to replicate our key findings from Study 1a even in this more consequential task.

3.1. Method

3.1.1. Participants

One hundred adults (50% females, M = 29.98 years, SD = 13.50) from the general population of a large Midwest city participated in a short in-laboratory study. They were paid \$2 for the study and received additional payoffs based on the condition to which they were assigned.

3.1.2. Design and procedure

Participants were randomly assigned to either the positive or negative reciprocity condition. Participants were told, "You will read a short description about a decision that someone made earlier in our laboratory. You will be asked to make some decisions and evaluations. You have already made \$2 for this survey and based on decisions of other players you will be given some additional money as a bonus." Participants were then told about an economic game (similar to Study 1a) in which they were asked to distribute resources to others. They read the following [within each set of brackets, the positive reciprocity condition appears first and the negative reciprocity condition appears second]:

Four people (you and three other people we will call Jake, Laura, and David) came into the lab and were asked to make decisions about how additional bonus payments should be divided. None of these people knew each other before today and these payments are on top of the \$2 each of you received as a show up payment.

(Jake, You, Laura, and David) have been asked to participate in this decision study. You will get to make decisions about how resources should be divided. You can divide the resources how you want among the two recipients.

In the first round, Jake was given \$4 to divide among you and Laura. [He gave \$3 to you and \$1 to Laura/ He gave \$1 to you and \$3 to Laura.]

In the second round, you have been given \$4 to divide among Jake and David. There are only two rounds and all four people now know what happened in round 1.

Participants then made their choice. Their options were to "give \$2 to Jake and \$2 to David," "\$3 to Jake and \$1 to David," and "\$1 to Jake and \$3 to David." After making

this decision, participants were asked three questions to evaluate the fairness of each of the three allocation options. For each option, the corresponding fairness measure read: "Please indicate your agreement or disagreement with the following: That would be fair." Participants indicated their agreement or disagreement on a 7-point Likert scale from 1 (*Strongly disagree*) to 7 (*Strongly agree*). Note, here we switched the order of the fairness and choice measures in order to avoid any influence of fairness considerations on participant choice.

Importantly, being assigned to these different conditions actually influenced participants' outcomes: In the positive reciprocity condition, participants received \$3 extra while in the negative reciprocity condition they only received \$1 extra. To make sure that they understood that this was consequential, they were given a special code and told to collect the additional money based on this code after the experiment.¹

3.2. Results

3.2.1. Fairness

A 2 (valence: positive, negative) × 3 (response: equality, reciprocity, or anti-reciprocity) mixed ANOVA with valence as a between-subjects factor and response as a within-subjects factor revealed no main effect of valence, F(1, 98) = .322, p = .572, $\eta_p^2 = .003$. However, we found a main effect of response, F(2, 97) = 40.41, p < .001, $\eta_p^2 = .292$, such that anti-reciprocity (M = 2.50, SD = 1.59) was seen as less fair than equal distribution (M = 4.59, SD = 1.84), t(99) = 9.07, p < .001, d = 1.22, and reciprocity (M = 4.30, SD = 1.87), t(99) = 7.28, p < .001, d = 1.04. Equality was not seen as more fair than reciprocity, t(99) = 0.99, p = .326, d = .16. Valence and response also interacted significantly, F(2, 97) = 7.55, p < .001, $\eta_p^2 = .071$, which we followed up on with planned contrasts. Fig. 1 includes the means and standard errors for the evaluations of reciprocity and equality.

Replicating Study 1a's results, negative reciprocity (M = 4.88, SD = 1.77) was perceived as more fair than positive reciprocity (M = 3.69, SD = 1.78), t(98) = 3.34, p = .001, d = .67. That is, people perceived the same inequality as more fair when it resulted from retaliation (negative reciprocity) rather than favoritism (positive reciprocity). Also participants' fairness ratings were not significantly different in their evaluations of equality in the negative (M = 4.51, SD = 1.77) and positive (M = 4.67, SD = 1.92) conditions, t(98) = .433, p = .666, d = .09. That is, equal distributions were not perceived differently regardless of previous history of positive or negative treatment.

Further, positive reciprocity was seen as less fair than equal division in response to positive treatment, t(48) = 2.60, p = .012, d = .75, whereas negative reciprocity was viewed as similarly fair to equal treatment in response to negative treatment, t (50) = 0.86, p = .392, d = .24.

We also found that anti-reciprocity was seen as more fair in the positive (M = 2.86, SD = 1.79) than negative condition (M = 2.16, SD = 1.30), t(87.46) = 2.23, p = .028, d = .45. We note this was not an a priori prediction and so we do not discuss it at length below, though we return to the issue of anti-reciprocity in the General Discussion.

10 of 35

3.2.2. Choice

We then analyzed people's choices to examine whether people engaged in reciprocity, equal allocation, or anti-reciprocity. We first conducted a planned contrast Fisher's exact test comparing people's tendency to choose reciprocity vs. equal allocation in the positive and negative conditions. For these analyses, we ignore anti-reciprocity. This information is included in the goodness of fit tests below, although it should be noted that this option was chosen infrequently.

When examining reciprocity and equality responses, we found that people were much more likely to say they would reciprocate in the negative (53% reciprocity vs. 47% equal) than the positive conditions (22% reciprocity vs. 78% equal), p < .001. Note that in this case, the choice measures came before the evaluations and yet we still observed the same pattern that we found in Study 1a, which should assuage any concerns that the choice results from Study 1a were driven by the fact that people made fairness evaluations before choosing how to respond.

We further conducted chi-squared goodness of fit tests. In the negative reciprocity conditions, participants picked reciprocity (51%) and equal division (45%) more often than anti-reciprocity (4%), $\chi^2(2, N = 51) = 20.02$, p < .001. In the positive conditions, they picked equal giving (73.5%) more often than reciprocity (20.5%) or anti-reciprocity (6%), $\chi^2(2, N = 49) = 37.02$, p < .001.

We also examined the correlation between people's evaluation of reciprocity and their choice (again this excluded the few participants who engaged in anti-reciprocity for simplicity of analyses). Participants' choices were correlated with their evaluation of reciprocity, r(95) = .62, p < .001, and equality, r(95) = -0.55, p < .001. This result demonstrates a correspondence between people's fairness evaluations and their choices—for example, if they had chosen equality, then they were more likely to think that equality was fair. From these data, we cannot make a causal conclusion about whether fairness evaluations drove choice or vice versa.

3.3. Discussion

As predicted and in line with the results from Study 1a, participants thought that it was more fair to engage in negative reciprocity than positive reciprocity. Furthermore, participants generally thought that it was more fair to distribute equally than to engage in reciprocity.

Our data also reveal that people not only show differential evaluations of the two forms of reciprocity, but that they themselves are also more willing to engage in negative reciprocity than positive reciprocity in multilateral interactions. This tendency to engage more in negative reciprocity than positive reciprocity is similar to findings from research on reciprocity in dyadic interactions (Charness & Rabin, 2002; Gray, Ward, & Norton, 2014; Keysar, Converse, Wang, & Epley, 2008; Offerman, 2002). While people may engage in negative reciprocity more than positive reciprocity in both types of interactions, we argue that negative reciprocity should be seen as more fair than positive reciprocity in multilateral interactions only. In another study (Supplemental Study S1), we found that negative reciprocity is not seen as more fair than positive reciprocity.

4. Study 2

In Study 2, we investigated judgments of resource distribution in a different multilateral context to examine if our effects from Studies 1a and 1b would generalize to a context in which a government official must decide how to allocate resources between different special interest groups. Participants read about Smith and Adams who were candidates in a senatorial election. The winner of the election then had the opportunity to divide government resources between two similar special interest groups, one of which paid money to get Smith elected. In one condition, Smith won and in the other Adams won. The winning senator then divides resources between the group, either equally (equality conditions) or unequally by paying back the group for their actions during the election (reciprocity conditions). In the positive reciprocity condition, Senator Smith gave more to the special interest group who had supported him. In the negative reciprocity condition, Senator Adams gave them less because they had supported Senator Smith. In all conditions, participants were asked to provide a fairness evaluation of the winning senator's actions and their likelihood of voting for that senator in a future election. The predictions were the same as in previous studies equality should be seen as more fair than reciprocity and negative reciprocity should be seen as more fair than positive reciprocity.

4.1. Method

4.1.1. Participants

Two hundred and one adults (46% females, M = 36.04 years, SD = 11.40) participated in this study.

4.1.2. Design and procedure

Participants were randomly assigned in a 2 (valence: positive, negative) \times 2 (response: reciprocity, equality) between-subjects design. Participants were told, "Please read about the following vignette and make your evaluation." All participants read the following:

James is a government lobbyist who works for an agriculture special interest group.

In the most recent state elections James donated a large amount of money, on behalf of the agriculture special interest group he works for, to Senator Smith's election campaign against Senator Adams. James's special interest group also ran a television ad campaign in support of Senator Smith so their support for Senator Smith is wellknown.

Participants then read that the supported senator won (positive condition) or that the non-supported senator won (negative condition) and also how that senator decided to allocate between the two special interest groups (within each set of brackets the reciprocity

condition appears first and the equality condition appears second). Participants in the positive conditions read:

Senator Smith won the election and is in charge of the subcommittee that will allocate subsidies between two large agricultural groups in his state: One of these groups is the group that James works for and the other is another agriculture interest group.

Senator Smith has a fixed amount of subsidies that he can assign between the two deserving groups and has the following three options:

- 1. Give most of the subsidies to James's special interest group
- 2. Give most of the subsidies to the other agricultural special interest group
- 3. Give an equal amount of subsidies to both groups.

Senator Smith decides to [give more subsidies to James's special interests group/provide equal subsidies to both agricultural special interests groups].

While participants in the negative conditions read:

Senator Adams won the election and is in charge of the subcommittee that will allocate subsidies between two large agricultural groups in his state: One of these groups is the group that James works for and the other is another agriculture interest group.

Senator Adams has a fixed amount of subsidies that he can assign between the two deserving groups and has the following three options:

- 1. Give most of the subsidies to James's special interest group
- 2. Give most of the subsidies to the other agricultural special interests group
- 3. Give an equal amount of subsidies to both groups

Senator Adams decides to [provide more subsidies to the other agricultural special interests group/provide equal subsidies to both agricultural special interests groups].

Participants then rated their agreement with two statements about the senator's actions, "I agree with what the senator did" and "What the senator did was fair," on a 7-point Likert scale that went from 1 (*Strongly disagree*) to 7 (*Strongly agree*). These items were presented in a random order and were highly correlated, r(201) = 0.87, p < .001, so we combined them and report a "general fairness" measure. We also collected a measure of

their likelihood of voting for the senator for reelection, which said, "Given what happened, and if I were a constituent of Senator Smith/Adams (depending on condition), I would vote for him in the next election," which was also presented on a 7-point Likert scale from 1 (*Disagree*) to 7 (*Agree*) scale.

4.2. Results and discussion

4.2.1. Fairness

We first conducted a 2 (valence: positive, negative) × 2 (response: reciprocity, equality) ANOVA that revealed a main effect of valence, F(1, 197) = 9.07, p = .003, $\eta_p^2 = .044$. This result indicates that participants thought that the senator was more fair in the negative (M = 5.02, SD = 1.62) than positive conditions (M = 4.37, SD = 2.09). We also found a main effect of response, F(1, 197) = 143.32, p < .001, $\eta_p^2 = .421$, such that participants saw the senator's actions as more fair in the equality conditions (M = 5.91, SD = 1.29) than in the reciprocity conditions (M = 3.52, SD = 1.62). Further, there was a significant valence by response interaction, F(1, 197) = 8.03, p = .005, $\eta_p^2 = .039$, which we followed up on with planned comparisons.

We compared fairness ratings for the different valence levels first in the reciprocity conditions, then in the equality conditions. As predicted, negative reciprocity (M = 4.11, SD = 1.57) was perceived as more fair than positive reciprocity (M = 2.94, SD = 1.46), t (100) = 3.88, p < .001, d = .77. That is, participants thought the same inequality was more fair when it was based on negative reciprocity rather than positive reciprocity. Further, participants in the equality conditions rated the senator's actions as similarly fair whether in the negative (M = 5.93, SD = 1.22) or positive (M = 5.90, SD = 1.38) condition, t(97) = 0.14, p = .892, d = .02. Fig. 2 shows the interaction results—fairness ratings only differed across valence in the reciprocity conditions.

As in Study 1a and 1b, participants in Study 2 thought that it was more fair to engage in negative than positive reciprocity. Furthermore, participants generally thought that it was more fair to distribute equally than engage in either form of reciprocity.

4.2.2. Voting decision

We next conducted a 2 (valence: positive, negative) × 2 (response: reciprocity, equality) ANOVA on the voting measure, which revealed a main effect of valence, F(1, 197) = 6.20, p = .014, $\eta_p^2 = .030$. This result indicates that participants were more likely to vote for the senator for reelection in the negative conditions (M = 4.78, SD = 1.53) than the positive conditions (M = 4.16, SD = 1.35). We also found a main effect of response, F(1, 197) = 143.32, p < .001, $\eta_p^2 = .241$, such that participants were more likely to vote for the senator in the equality conditions (M = 5.33, SD = 1.46) than in the reciprocity conditions (M = 3.74, SD = 1.45). There was also a valence and response interaction, F(1, 197) = 8.03, p = .005, $\eta_p^2 = .014$, which we followed up on with planned comparisons.

We compared participants' likelihood of voting for the different valence levels first in the reciprocity conditions, then equality conditions. As predicted, people were more likely

14 of 35



Fig. 2. Mean fairness rating by condition in Study 1b. Bars represent SE's of the mean.

to vote for the candidate who engaged in negative (M = 4.16, SD = 1.35) rather than positive (M = 3.31, SD = 1.44) reciprocity, t(100) = 3.06, p = .003, d = .61. That is, although the winning senator's actions created the same inequality, participants were more likely to re-elect the candidate who committed negative reciprocity rather than the one who participated in positive reciprocity. Further, participants in the equality conditions did not respond differently to whether they would re-elect the senator in the negative (M = 5.42, SD = 1.45) or positive (M = 5.25, SD = 1.49) conditions, t(97) = 0.55, p = .585, d = .12. That is, the pattern of participants' self-reported willingness to vote for the senator corresponded to their fairness evaluations.

Thus, in a scenario involving decision-making in a governmental context, we replicate our results from Study 1a and 1b, finding that equality is seen as more fair than reciprocity and that negative reciprocity is seen as more fair than positive reciprocity.

5. Study 3

We have argued that negative reciprocity is seen as more fair than positive reciprocity because the former does a better job of mitigating the negative inferences of partiality and favoritism that normally result from unequal treatment. In Study 3, we explored this mechanism directly by measuring people's beliefs about whether the inequality occurred because of a pre-existing preference and whether this favoritism will occur in the future. We did this by replicating the positive and negative reciprocity conditions from Study 2, but adding measures of perceived favoritism. If people's differential fairness evaluations are being driven by the role of perceived favoritism in this context, then positive reciprocity should be perceived as being more clearly rooted in favoritism than negative reciprocity and this difference in favoritism should mediate the difference in fairness evaluations.

16 of 35

5.1. Method

5.1.1. Participants

One hundred adults (48% females, M = 37.64 years, SD = 12.33) participated in this 5-min study for \$0.25.

5.1.2. Procedure

The procedure was an exact replication of the positive and negative reciprocity conditions from Study 2, except that we added a favoritism measure. After reading the scenario, participants answered two fairness questions identical to those in previous studies. These items were again highly correlated, r(100) = 0.86, p < .001, and combined into one "general fairness" measure. Participants were also asked two new questions to measure favoritism on a 7-point Likert scale ranging from 1 (Strongly disagree) to 7 (Strongly agree) scale. The questions asked participants to rate their agreement with the following statements: "[Senator Smith/Senator Adams] has a pre-existing preference for [the special interests group James works for/the other special interests group (not the one James works for)]" and "[Senator Smith/Senator Adams] will show favoritism toward [the special interests group James works for/the other special interests group (not the one James works for)] in the future." The two measures were also correlated r(100) = 0.69, p < .001 so we combined these into a "favoritism" measure. Note, in all of our subsequent studies we combine these two measures, but we get similar results and mediation in all studies if we just use either measure. We randomized the order of the fairness and favoritism questions.

5.2. Results and discussion

5.2.1. Fairness

Participants rated the allocator as more fair in the negative (M = 4.02, SD = 1.34) than positive reciprocity condition, (M = 3.33, SD = 1.60), t(98) = 2.34, p = .022, d = .47. Fig. 3 includes the means and standard errors for all conditions. These replicate the main result from our previous studies.

5.2.2. Favoritism measure

The allocator in the negative reciprocity condition (M = 4.86, SD = 1.55) was seen as showing less favoritism than the allocator in the positive reciprocity condition, (M = 5.50, SD = 1.20), t(98) = 2.31, p = .023, d = .46.

5.2.3. Voting measure

Participants stated that they were more likely to vote for the senator in the negative (M = 3.84, SD = 1.51) than positive reciprocity condition, (M = 3.24, SD = 1.20), t (97) = 2.19, p = .031, d = .44.



Fig. 3. Mean fairness and favoritism rating by condition in Study 3, Supplemental Study S2, and Study 6. Bars are standard errors.

5.2.4. Mediation analysis

Finally, we tested whether favoritism mediated the effect of condition on fairness evaluations. We found that valence (positive vs. negative reciprocity) influenced reported fairness ($\beta = 0.69$, p = .022) and perceived favoritism ($\beta = -0.64$, p = .023). Perceived favoritism was related to fairness judgments ($\beta = -0.46$, p < .001), and the inclusion of perceived favoritism in the analysis reduced the effect of condition on fairness evaluations ($\beta = 0.39$, p = .157). A bootstrap analysis (10,000 bootstrapped sample) revealed that the 95% bias-corrected confidence interval for the size of the indirect effect excluded zero (0.04, 0.68), suggesting a significant indirect effect of favoritism (MacKinnon, Fairchild, & Fritz, 2007). Thus, as expected, perceived favoritism mediated the relationship between fairness evaluations in the positive and negative reciprocity conditions (Supplemental Study S2 includes a replication of these results in an office setting).

6. Study 4

In Study 4, we compared fairness evaluations of positive and negative reciprocity to the most common and acceptable justification for unequal distribution of resources: merit (Adams, 1965). At least in Western societies, people willingly accept that those who contribute more or perform better should receive more, and inequality based on merit is not considered a form of favoritism (e.g., Adams, 1965; Hook & Cook, 1979). Merit, then, provides a useful comparison point between people's evaluations of inequality due to a highly impartial justification and justification based on the two types of reciprocity. Given that merit-based inequality involves no favoritism, we expected that unequal allocation based on merit should be seen as most fair followed by negative reciprocity, which would

be seen as more fair than positive reciprocity. Furthermore, we again expected perceptions of favoritism to mediate these differences in fairness evaluations.

6.1. Method

6.1.1. Participants

One hundred and fifty-one adults (44% females, M = 38.68 years, SD = 11.54) participated in this study.

6.1.2. Procedure

The procedure was similar to our previous studies, but this time the scenario was set in an office and included new measures of how the participant would feel as an employee of the company. It was a hypothetical scenario about four co-workers who participants were asked to imagine that they worked with. The scenario read as follows (within each set of brackets the positive reciprocity condition appears first and the negative reciprocity condition appears second):

Imagine that you work at a large firm with James and Sally.

James was asked to divide 4 concert tickets purchased by the office between Sally and Rachel, two co-workers who work equally hard at their jobs and receive equally positive evaluations. James decides to give [Sally 3 tickets and Rachel 1 ticket/Sally 1 ticket and Rachel 3 tickets].

The following week, Sally, one of James' co-workers, was asked to decide how to assign a bonus among two employees, James and Bill, who both work equally hard at their jobs and receive equally positive evaluations. Sally could assign a total of a \$600 bonus between the two employees.

She could have assigned it equally, but instead she gave \$400 to James and \$200 to Bill.

In the merit condition participants read the following:

Imagine that you work at a large firm with James and Sally.

Sally, one of James' co-workers, was asked to decide how to assign a bonus among two employees, James and Bill. Although both employees work hard, James consistently receives better evaluations than Bill. Sally could assign a total of a \$600 bonus between the two employees.

She could have assigned it equally, but instead she gave \$400 to James and \$200 to Bill.

After reading one of these scenarios, participants in all conditions evaluated the fairness and favoritism demonstrated by Sally's unequal distribution. The fairness items were again highly correlated, r(152) = 0.92, p < .001 as were the favoritism measures, r(152) = 0.76, p < .001, so we combined them into a "fairness" and "favoritism" measure, respectively. We also added a measure to assess how participants would feel about the allocator being promoted to be their boss. They could agree or disagree with the following: "I would be satisfied if Sally were promoted to be my supervisor" on a 7-point Likert scale ranging from 1 (*Strongly disagree*) to 7 (*Strongly agree*).

6.2. Results and discussion

6.2.1. Fairness

A one-way ANOVA revealed a main effect of justification, F(2, 148) = 39.41, p < .001, $\eta_p^2 = .348$. We followed up on this with planned contrasts. The merit condition (M = 5.32, SD = 1.58) was seen as more fair than both the negative reciprocity condition (M = 3.60, SD = 1.67), t(96) = 5.26, p < .001, d = 1.06, and the positive reciprocity condition (M = 2.52, SD = 1.62), t(102) = 8.95, p < .001, d = 1.75. And the negative reciprocity condition, t (98) = 3.27, p = .001, d = .66. Fig. 4 includes the means and standard errors for all conditions.

6.2.2. Favoritism measure

A one-way ANOVA on the favoritism measure revealed a main effect of justification, F (2, 148) = 41.62, p < .001, $\eta_p^2 = .360$. We followed up on this with planned contrast t tests. The merit condition (M = 3.56, SD = 1.58) was seen as showing less favoritism than both the negative reciprocity condition (M = 4.67, SD = 1.22), t(96) = 3.87, p < .001, d = .79 and the positive reciprocity condition (M = 5.89, SD = 1.05), t (102) = 8.88, p < .001, d = 1.74. And the negative reciprocity condition was seen as showing less favoritism than the positive reciprocity condition, t(98) = 5.36, p < .001, d = 1.07. Fig. 4 includes the means and standard errors for all conditions.

6.2.3. Satisfaction measure

A one-way ANOVA on the satisfaction measure revealed a main effect of justification, F (2, 148) = 36.65, p < .001, $\eta_p^2 = .331$, which we followed up on with planned contrast t tests. Participants in merit condition (M = 4.84, SD = 1.57) were more satisfied with the allocator being their boss than those in both the negative reciprocity condition (M = 3.19, SD = 1.65), t(96) = 5.09, p < .001, d = 1.02, and the positive reciprocity condition (M = 3.19, SD = 1.65), t(96) = 5.09, p < .001, d = 1.69. We also found that participants in the negative reciprocity condition reported they would be more satisfied with the allocator as a boss than those in the positive reciprocity condition, t(98) = 3.09, p = .003, d = .61. This hypothetical measure accords with real-world negative reactions to favoritism (e.g., Cropanzano, Howes, Grandey, & Toth, 1997; Khatri & Tsang, 2003; Poon, 2004).

20 of 35

6.2.4. Mediation analysis

Finally, we tested whether favoritism mediated the effect of justification on fairness evaluation. First, we again examined positive and negative reciprocity. We found that justification (positive reciprocity vs. negative reciprocity) influenced reported fairness ($\beta = 1.08$, p = .002) and perceived favoritism ($\beta = -1.22$, p < .001). Perceived favoritism was related to fairness judgments ($\beta = -0.32$, p = .03), and the inclusion of perceived favoritism in the analysis reduced the effect of justification on fairness evaluations ($\beta = .69$, p = .06). A bootstrap analysis (10,000 bootstrapped sample) revealed that the 95% bias-corrected confidence interval for the size of the indirect effect excluded zero (0.11, 0.81). Thus, as expected, perceived favoritism mediated the relationship between fairness evaluations in the positive and negative reciprocity conditions.

Next, we tested whether favoritism mediated the effect of justification on fairness evaluation for merit and negative reciprocity. We found that justification (negative reciprocity vs. merit) influenced reported fairness ($\beta = 0.86$, p < .001) and perceived favoritism ($\beta = -0.56$, p < .001). Reported favoritism was related to fairness judgments ($\beta = -0.33$, p = .004), and the inclusion of perceived favoritism in the analysis reduced the effect of justification on fairness evaluations ($\beta = 0.68$, p < .001). A bootstrap analysis (10,000 bootstrapped sample) revealed that the 95% bias-corrected confidence interval for the size of the indirect effect excluded zero (0.04, 0.37). Thus, as expected, perceived favoritism mediated the relationship between fairness evaluations in the merit and negative reciprocity conditions.

Finally, we tested whether favoritism mediated the effect of justification on fairness evaluations for merit and positive reciprocity. We found that justification (positive reciprocity vs. merit) influenced reported fairness ($\beta = .93$, p < .001) and perceived favoritism ($\beta = -0.78$, p < .001). Perceived favoritism was related to fairness judgments ($\beta = -0.50$, p < .001), and the inclusion of perceived favoritism in the analysis reduced the effect of justification on fairness evaluations ($\beta = .55$, p < .001). A bootstrap analysis



Fig. 4. Mean fairness rating and favoritism rating by condition in Study 4. Bars represent SE's of the mean.

) 21 of 35

(10,000 bootstrapped sample) revealed that the 95% bias-corrected confidence interval for the size of the indirect effect again excluded zero (0.17, 0.60). Thus, as expected, perceived favoritism mediated the relationship between fairness evaluations in the positive reciprocity and merit conditions.

Our results demonstrate that negative reciprocity does function as a better justification for unequal allocations than positive reciprocity, but that it is a weaker justification for inequality than merit. These differences between people's fairness evaluations were mediated by people's perceptions of favoritism—the allocator actions that were seen as most indicative of favoritism were thought to be the least fair.

One question is whether positive reciprocity is better or worse than inequality that has no justification. Importantly, in two supplemental studies we find that people think unequal treatment for no reason is even less fair than positive reciprocity (Supplemental Study S3a and S3b). As such, both positive and negative reciprocity appear to provide some justification for unequal division, but the latter does a better job than the former at mitigating inferences of favoritism and is therefore seen as more fair.

7. Study 5

We have shown that perceived favoritism mediates the finding that negative reciprocity is seen as more fair than positive reciprocity. If this is true, then this relationship between (un)fairness and favoritism should be true only for contexts where favoritism is viewed negatively. Study 5 investigates this hypothesis by comparing an office context where favoritism is indeed negative to a personal context where people distribute their own resources among friends. Based on previous research, we know that people react less negatively to partiality in interpersonal contexts (Everett, Faber, Savulescu, & Crockett, 2018; Fiske, 1992). Indeed, friends may expect a degree of favoritism from their friends in interpersonal interactions (Shaw et al., 2017). Thus, we expected that people would still see both positive and negative reciprocity as showing favoritism when someone was asked to divide personal resources between friends, but that this favoritism would not necessarily be seen as unfair.

7.1. Method

7.1.1. Participants

Two hundred adults (44% females, M = 33.04 years, SD = 10.28) participated in this study.

7.1.2. Procedure & design

We used a 2 (valence: positive, negative) \times 2 (context: office, personal) between-subjects design. We used vignettes similar to our previous studies, but we manipulated whether resources belonged to the office or were personal. In the office conditions, participants read (within each set of brackets the positive reciprocity condition appears first and the negative reciprocity condition appears second):

James works at an advertising firm.

Recently at work Sally, one of James' co-workers, was asked to decide how to assign tickets for a 'Disney on Ice' show among two employees, James and Bill. James and Bill both work equally hard at their jobs and receive equally positive evaluations.

The company has a total of 6 tickets to give and Sally has the following three options:

- 1. Give 4 tickets to James, and 2 tickets to Bill.
- 2. Give 3 tickets to James, and 3 tickets to Bill.
- 3. Give 2 tickets to James, and 4 tickets to Bill.

Sally decides to give 4 tickets to James, and 2 tickets to Bill.

Later at work, James was asked to divide 4 'Hamilton' tickets between Sally and Rachel, two equally hardworking co-workers.

James decides to give [Sally 3 tickets and Rachel 1 ticket/ Sally 1 ticket and Rachel 3 tickets].

The personal resource context was similar, but in this case, the resource belonged to the individual distributor (within each set of brackets the positive reciprocity condition appears first and the negative reciprocity condition appears second):

James works at an advertising firm and often spends time with his co-workers outside of work.

While hanging outside of work, Sally, one of James' co-workers, told James and a few others that she had received tickets for a 'Disney on Ice' show for her birthday recently but could no longer make it. So, Sally decides to give the tickets away to two fellow employees, James and Bill.

Sally has a total of 6 tickets to give and has the following three options:

- 1. Give 4 tickets to James, and 2 tickets to Bill.
- 2. Give 3 tickets to James, and 3 tickets to Bill.
- 3. Give 2 tickets to James, and 4 tickets to Bill.

22 of 35

Sally decides to give 4 tickets to James, and 2 tickets to Bill.

Later that month, James is hanging out with his co-workers again and tells them that he won 4 'Hamilton' tickets recently in a raffle. James tells his co-workers that he can't make the show and decides to split the tickets between Sally and Rachel, two of his co-workers.

James decides to give [Sally 3 tickets and Rachel 1 ticket/ Sally 1 ticket and Rachel 3 tickets].

Participants were asked the same fairness (two-item correlation: r(200) = .80, p < .001) and favoritism questions (two-item correlation: r(200) = .65, p < .001) as in our previous studies.

7.2. Results

7.2.1. Fairness

We first conducted a 2 (valence: positive, negative) × 2 (context: office, personal) ANOVA on the fairness measure, which revealed no main effect of valence, F(1, 196) = 0.30, p = .587, $\eta_p^2 = .002$. We found a main effect of context, F(1, 196) = 37.14, p < .001, $\eta_p^2 = .159$, such that participants saw reciprocity as less fair in the office condition (M = 3.27, SD = 1.72) than in the personal condition (M = 4.62, SD = 1.57). There was also a significant valence by context interaction, F(1, 196) = 16.62, p < .001, $\eta_p^2 = .078$, which we followed up on with planned contrasts.

In the office conditions, we found that negative reciprocity was seen as more fair (M = 3.65, SD = 1.61) than positive reciprocity (M = 2.88, SD = 1.59), t(100) = 2.44, p = .016, d = .49. The opposite was true in the personal context where positive reciprocity (M = 5.07, SD = 1.51) was seen as more fair than negative reciprocity (M = 4.12, SD = 1.51), t(96) = 3.11, p = .002, d = -.63. Fig. 5 includes the means and standard errors for all conditions.

7.2.2. Favoritism measure

We conducted a 2 (valence: positive, negative) × 2 (context: office, personal) ANOVA on the favoritism measure, which revealed a main effect of valence, F(1, 196) = 45.02, p < .001, $\eta_p^2 = .187$, such that positive reciprocity (M = 5.45, SD = 1.30) demonstrated more favoritism than negative reciprocity (M = 4.28, SD = 1.30). We also found a main effect of context, F(1, 196) = 3.15, p = .078, $\eta_p^2 = .016$, with participants reporting reciprocity as showing marginally more favoritism in the office condition (M = 5.00, SD = 1.30) than in the personal condition (M = 4.73, SD = 1.43). There was no valence by context interaction, F(1, 196) = .795, p = .374, $\eta_p^2 = .004$.

7.2.3. Mediation

In the office context, valence influenced reported fairness ($\beta = .77$, p = .016) and perceived favoritism ($\beta = -1.32$, p = .005). Perceived favoritism was related to fairness judgments ($\beta = -0.26$, p = .026), and the inclusion of perceived favoritism in the analysis reduced the effect of condition on fairness evaluations ($\beta = .25$, p = .485). A bootstrap analysis (10,000 bootstrapped sample) revealed that the 95% bias-corrected confidence interval for the size of the indirect effect again excluded zero (0.06, 1.07). As expected, perceived favoritism mediated the relationship between fairness evaluations in the positive and negative reciprocity conditions.

In the personal context, valence influenced reported fairness ($\beta = -1.01$, p = .001) and perceived favoritism ($\beta = -1.01$, p < .001). Perceived favoritism, however, did not influence fairness judgments significantly ($\beta = -0.12$, p = .294). Thus, we could not conduct a mediation analysis because one of the assumptions of mediation was not met.

We replicated our previous results from Study 4 when someone was dividing office resource, but found that when people were dividing their own resources they responded quite differently. Here, positive reciprocity was seen as more fair than negative reciprocity. Interestingly, in both the office and interpersonal contexts, we found that positive reciprocity is seen as showing greater favoritism than negative reciprocity. However, perceptions of favoritism were correlated with fairness judgments only when dividing office resources, not when dividing personal resources. In the interpersonal realm, people thought that the favoritism demanded by positive reciprocity was fair. Indeed, people often expect and desire partiality from their friends (DeScioli & Kurzban, 2009a; Everett et al., 2018; Hughes, 2017; Shaw et al., 2017).

8. Study 6

We have argued that perceived favoritism explains our findings but another reason that people may be less accepting of positive reciprocity is that it negatively impacts someone who did nothing wrong. Most of our previous experiments cannot adjudicate between these two accounts because they involved a zero-sum resource distribution—one recipient's gain was another's loss.

In order to evaluate this alternative account, Study 6 examined situations of non-zerosum resource distribution. In this case, the distributor can either give an additional resource to a particular recipient or give the resource to no one (Choshen-Hillel et al., 2015; Shaw & Knobe, 2013). Thus, giving extra to one party does not reduce what the other receives. If the alternative account is correct, then there should be no difference in evaluations of positive and negative reciprocity in this non-zero-sum situation because there is no consequence for the other recipient. However, if positive reciprocity is still perceived as more unfair than negative reciprocity, this would suggest that our findings are not due to solely concerns about collateral damage to the third party.

8.1. Method

8.1.1. Participants

One hundred adults (51% females, M = 34.9 years, SD = 9.90) participated in this study.

24 of 35



Fig. 5. Mean fairness rating and favoritism rating by condition in Study 5. Bars represent SE's of the mean.

8.1.2. Procedure

We used vignettes similar to Study 4, non-merit conditions. The scenario was also similar, except that now reciprocity did not affect the other recipient because the distributor only had two options: to distribute unequally by giving the resource to one recipient or give the resource to neither participant. The scenario read as follows (within each set of brackets, the positive reciprocity condition appears first and the negative reciprocity condition appears second):

Imagine that you work at the firm with James and Sally.

James was asked to divide 4 concert tickets purchased by the office between Sally and Rachel, two co-workers who work equally hard at their jobs and receive equally positive evaluations. James decides to give [Sally 3 tickets and Rachel 1 ticket/Sally 1 ticket and Rachel 3 tickets].

The following week, Sally was asked to decide how to assign a bonus among two employees, James and Bill, who both work equally hard at their jobs and receive equally positive evaluations. The company can give a total of a \$500 bonus but based on institutional rules the bonuses must be in \$100 increments. Sally has given each employee a \$200 bonus. She now must decide what to do with the remaining \$100 bonus. Neither employee will know what the other one receives.

Positive: She can either give the remaining \$100 to James or give it to neither employee. She decided to give it to James so he received a \$300 bonus and Bill received a \$200 bonus.

Negative: She can either give the remaining \$100 to Bill or give it to neither employee. She decided to give it to Bill so he received a \$300 bonus and James received a \$200 bonus.

Participants were thus told that James initially engaged in unequal distribution and that subsequently Sally engaged in reciprocity, but in this case, the reciprocity did not come at a cost to the other co-worker because the only options available were to give the bonus to one person or to neither. To further clarify this point, in the negative condition Sally's only options were to give it to Bill or give it to neither, and thus, giving it to Bill did not take anything away from James.

Participants were given the same questions as in Study 4. The two fairness measures were again highly correlated, r(100) = 0.72, p < .001, as were the favoritism questions, r(100) = 0.70, p < .001, and therefore combined into one fairness and one favoritism measure, respectively.

8.2. Results and discussion

8.2.1. Fairness

An independent samples t test revealed that negative reciprocity (M = 4.59 SD = 1.16) was seen as more fair than positive reciprocity (M = 3.66, SD = 1.49), t(98) = 3.49, p = .001, d = .70.

8.2.2. Favoritism measure

An independent samples t test revealed that negative reciprocity (M = 4.48, SD = 1.31) was seen as showing less favoritism than positive reciprocity (M = 5.20, SD = 1.40), t(98) = 2.64, p = .010, d = .53. Fig. 3 includes the means and standard errors for all conditions.

8.2.3. Mediation analysis

Finally, we tested whether favoritism mediated the effect of condition on fairness evaluations for positive and negative reciprocity. We found that valence condition (positive vs. negative reciprocity) influenced reported fairness ($\beta = .93$, p < .001) and perceived favoritism ($\beta = -0.72$, p < .010). Perceived favoritism was related to fairness judgments ($\beta = -0.24$, p = .014), and the inclusion of perceived favoritism in the analysis reduced the effect of the valence condition on fairness evaluations ($\beta = .76$, p = .006). A bootstrap analysis (10,000 bootstrapped sample) revealed that the 95% bias-corrected confidence interval for the size of the indirect effect excluded zero (0.02, 0.45). Thus, perceived favoritism again mediated the relationship between fairness evaluations in the positive and negative reciprocity conditions.

We thus replicated our basic findings in a non-zero-sum situation. This finding is important because it demonstrates that our previous results are not due solely to the fact that someone suffers as a result of the unequal distribution. In this study, the decision to give the resource entailed no cost to the other recipient because the decision-maker only had the option to give according to reciprocity or give to neither. We even made it clear that the other party would not know about the potential inequality to minimize concerns of negative social comparison being a potential harm (Festinger, 1954). Still, negative reciprocity was perceived as more fair than positive reciprocity even in a case where no one was harmed. Of course, these data do not rule out the idea that one of the reasons that people may think negative reciprocity is more fair is because it harms someone who was "bad." However, it does demonstrate that such harm is not necessary for people to see negative reciprocity as more fair than positive reciprocity.

9. General discussion

We found evidence for both of our hypotheses about how people evaluate resource allocation in multilateral interactions: Equal distribution was seen as more fair than reciprocity, and negative reciprocity was seen as more fair than positive reciprocity. We found support for these hypotheses in multilateral interactions involving economic games, government, and organizational decision-making situations. Table 1 provides a summary of these effects across our studies. Importantly, we found this pattern only in multilateral interactions when people are distributing resources that are not theirs, but not in situations where they are dividing their own resources between friends (Study 5), nor in dyadic interactions (Supplemental Study S1). We further found that this differential evaluation of positive and negative reciprocity is partly driven by people's inferences about partiality (Studies 3, 4, 5, and 6). Further, people were more likely to say that they would vote for a candidate who engaged in negative rather than positive reciprocity (Study 2 and 3), and believed they would be more satisfied working at a company where a potential supervisor engaged in negative rather than positive reciprocity (Study 4). Moreover, we found that merit provides the best justification for unequal outcomes (Study 4) and is viewed as the most fair unequal division strategy that we investigated. Importantly, we find that negative reciprocity is seen as more fair than positive reciprocity even when reciprocating has

Table 1

Summary	of	Mean	(SD)	fairness	evaluations	for	the	positive	and	negative	reciprocity	conditions	from	our	7
studies															

	Positive Reciprocity	Negative Reciprocity	N	t	d
Study 1a (Econ Game)	3.91 (1.65)	5.22 (1.55)	84	3.73***	.82
Study 1b (Econ Game)	3.69 (1.78)	4.88 (1.77)	100	3.34***	.67
Study 2 (Government)	2.94 (1.65)	4.11 (1.46)	102	3.88***	.75
Study 3 (Government)	3.33 (1.34)	4.02 (1.91)	100	2.34^{*}	.42
Study 4 (Office)	2.52 (1.62)	3.60 (1.67)	100	3.27***	.66
Study 5 (Office)	2.88 (1.59)	3.65 (1.61)	102	2.44^{*}	.48
Study 6 (Office)	3.66 (1.52)	4.59 (1.53)	100	3.59***	.61

Note The parenthesis next to the study name refers to the situation involved. Higher numbers equal higher fairness evaluations.

 $*p \le .05;$

 $**p \le .01;$

*** $p \le .001.$

no negative impact on someone else (Study 6). Our results reveal that negative reciprocity may be taken as a justification for inequality that excuses the partiality that is usually entailed when one person is favored over another.

These results provide further support to the partiality account of fairness (Shaw, 2013) and highlight the tension that exists between partiality and fairness. Indeed, recent research has suggested that "fairness" may not be needed to explain why people support redistribution of resources (e.g., Sznycer et al., 2017) or get upset at receiving less than others (Shaw & Choshen-Hillel, 2017; Shaw, Choshen-Hillel, & Caruso, 2016), but it might be needed to explain people's negative reaction to partiality. Many models about the evolution of cooperation and fairness seem to assume that being fair and being a good reciprocator are complementary goals (Baumard et al., 2013; Falk et al., 2008). Our results demonstrate that this is not the case. Fairness and reciprocity can often be at odds with one another. Indeed, notions of reciprocity may tap into our notions of loyalty and favoritism, which are fundamentally at odds with our notions of fairness (Dungan et al., 2014; Shaw et al., 2012; Waytz, Dungan, & Young, 2013). Furthermore, pre-existing alliances will likely license more negative inferences about positive reciprocity-people may be reluctant to give a friend at work a bonus because it could be perceived as showing favoritism (e.g., Shaw, Choshen-Hillel, & Caruso, 2018). Relatedly, while some argue that being "fair" is a good strategy for attracting potential trading partners and allies (Baumard et al., 2013), people often value friends and exchange partners who are biased in their favor (Barakzai & Shaw, 2018; Shaw et al., 2017; Waytz et al., 2013). Indeed, a friend who was always "fair" and never preferentially took your side would be an excellent judge, but a crummy friend (DeScioli & Kurzban, 2009b, 2011; Shaw, 2016). These studies emphasize the trade-offs that people must sometimes make between being a good cooperator with someone and being an impartial agent.

Our experiments join a small set of studies (Niemi et al., 2018; Niemi & Young, 2017) that have examined how people make fairness evaluations about equality and reciprocity in multilateral interactions and they demonstrate that the correspondence between fairness and reciprocity is much weaker in multilateral interactions. These multilateral interactions have fundamentally different dynamics, which is why we found a clear divergence in people's fairness evaluations between dyadic and multilateral interactions. Whereas people thought negative reciprocity in the service of retaliation was much more fair than favoritism through positive reciprocity in multilateral interactions, people did not think that negative reciprocity was more fair than positive reciprocity in dyadic interactions (Supplemental Study S1). This work accords with recent work (Weisel & Shalvi, 2015) highlighting the ways in which working together with others can sometimes lead to increased corruption rather than fairness. Indeed, our results suggest that, in some contexts, politicians and other decision-makers may be able to mitigate negative reactions to inequality by framing them as retaliation for a lack of support (Studies 1 and 3) rather than as a form of favoritism.

Of course, how people react to positive or negative reciprocity will likely depend on the inferences they make about the mental state of the person allocating resources. Here we found evidence that negative reciprocity was evaluated as more fair than positive reciprocity and argue that this occurred because of the underlying inferences about internal mental states of the distributor, specifically his or her perceived favoritism toward the agent (Studies 3-6). However, there are certainly cases where negative reciprocity will lead to more negative inferences about the distributor's mental state, particularly if the retaliation is cruel or leads to disproportionate punishment. There is reason to believe that retaliation often entails disproportionate payback (e.g., Keysar et al., 2008; Rubin, Pruitt, & Kim, 1994; Schroeder, Steel, Woodell, & Bembenek, 2003). Such disproportional punishment through negative reciprocity may lead to differential evaluations of fairness since previous research has demonstrated that people make fairness judgments partly based on proportionality (Carlsmith, 2006). A real-life case of disproportionate retaliation illustrates this point. It occurred when a staff member of Chris Christie, the governor of New Jersey, closed lanes of a bridge to Fort Lee as payback for Fort Lee's mayor refusing to endorse Christie during the 2013 gubernatorial elections. People decried this as unfair, presumably because it was a disproportionate punishment and also negatively impacted other innocent people. The subsequent fallout significantly hurt Christie's career and the public's perception of him (Kleinfeld, 2014). More broadly, future research might investigate the inferences that people make about the mental states of allocators as they engage in reciprocity as this will likely exert a dramatic impact on people's eventual judgments of the allocators.

The current studies did not systematically investigate how people evaluate "anti-reciprocity"-that is, paying back kindness with unkindness and vice versa. Part of the reason that we did not investigate this possibility is that we assumed it would be rarely chosen (which is what we found in Study 1a and 1b). However, it is interesting to think about how people interpret and evaluate such behavior. Anti-reciprocity in response to preferential treatment from someone might have a straightforward interpretation: The person may want to distance themselves from this act of favoritism or be seen as punishing the unfairness. However, anti-reciprocity in response to negative treatment would be more difficult to interpret. One possibility is that such an action will be seen as particularly kind (i.e., an extreme form of "turning the other cheek"). Or, it could be seen as a form of taking the moral high-ground; this possibility is demonstrated clearly with former First Lady Michelle Obama's now-famous phrase, "When they go low, we go high." A further possibility is that this may just seem odd and people may react negatively to someone being rewarded for being mean or unfair. In the one study where we examined participants' evaluations of anti-reciprocity (Study 1b), participants thought that anti-reciprocity in response to negative treatment was unfair and less fair than antireciprocity in response to positive treatment. However, we certainly think there may be cases where anti-reciprocity is interpreted positively; these questions are beyond the scope of this paper.

Our experiments demonstrated that people who do not benefit from partiality find partiality to be unfair, but clearly the people who benefit from favoritism are sometimes happy to continue such favoritism despite accusations of unfairness. Indeed, even in contexts where people think partiality is unfair, people still observe a high degree of favoritism (Reinsch & Gardner, 2014; Waytz et al., 2013). In some cases, such favoritism could propagate because the distributors and recipients accept the reputational costs of being unfair in favor of building a mutually beneficial alliance with specific employees. However, we also suspect that in many of these cases, people attempt to cloak their favoritism in plausibly deniable excuses for their unequal treatment. Indeed, people are masterful at being able to justify their actions (e.g., Hsee, 1996; Kunda, 1990; Shalvi, Gino, Barkan, & Ayal, 2015). Further, what people value as a fair decision rule can shift based on whether the rule is aligned with the distributor's motives (DeScioli, Massenkoff, Shaw, Petersen, & Kurzban, 2014; Messick & Sentis, 1979). We imagine similar processes play out in the workplace, with people engaging in reciprocity by using similar justifications to themselves and others to cloak the apparent favoritism.

Although most of our experiments focused on third party decision-making, contexts in which partiality and favoritism are generally regarded as unfair, there are of course many situations, even in Western cultures, in which people are not judged harshly for being partial (e.g., as in our Study 5). People often expect loyalty and preferential treatment in interpersonal relationships, and as a result, they judge others negatively for not preferentially helping loved ones (Everett et al., 2018; Hughes, 2017) and respond negatively to neutrality from friends (Shaw et al., 2017). We hope that our work joins the considerable work that has been done on how people navigate these demands in different relational contexts (e.g., Fiske, 1992).

Importantly, the emphasis placed on impartiality and the kind of justifications that mitigate these partiality inferences will likely vary by culture (Shaw, 2016). There are two ways that cultures may respond differently. First of all, some cultures may treat the workplace more like the interpersonal domain in which favoritism is acceptable and, in that case, we would not expect them to view positive reciprocity as unfair. While many cultures may agree that impartiality is important in some spheres, particularly in legal proceedings involving punishment, (DeScioli & Kurzban, 2009a; Shaw, 2016), they might have considerable disagreements about when impartiality is or is not important. Secondly, a culture may agree that favoritism is bad in a particular workplace context, but not see reciprocity or nepotism as forms of favoritism. In these societies, they may agree that favoritism is unfair in the workplace but disagree that reciprocity demonstrates favoritism (Arasli & Tumer, 2008; Boadi, 2000; Brody, Coulter, & Mihalek, 1998; Fiske & Rai, 2014; Henrich et al., 2010; Miller & Bersoff, 1992; Yamagishi, Jin, & Miller, 1998). In some of these societies people may see even positive reciprocity ("Of course it is not partial to give more to someone who gave more to me previously") the way that people in Western societies perceive merit ("Of course it is not partial to give more to the person who did a better job"). Indeed, in some collectivist cultures there is an even greater link between equality and reciprocity with the principle of renging in China and on in Japan. Both of these principles emphasize social relationships and speak toward maintaining harmony in relationships through reciprocity. Thus, it seems possible that reciprocity would be more likely to be seen as a legitimate rule for unequal distribution in such societies. How cultures establish the contexts in which favoritism is viewed as unfair and the acceptable justifications for inequality is an important open question for future research.

31 of 35

The current research explored how concerns with equality and reciprocity interact in multilateral situations and demonstrates that reciprocity is often viewed as unfair because reciprocity displays unequal and partial treatment of others. We fully acknowledge that the desire to repay favors and to show favoritism toward one's friends is a powerful force in both the United States and across the globe. It is precisely because of this desire to be partial toward those who are close to us that, in some cultures, people disapprove of the appearance of favoritism in situations involving power hierarchies and the fruits of group labor, such as at the workplace, in government, and in criminal justice. Many public policy debates about the fairness of a decision involve tradeoffs between reciprocity and fairness. Therefore, it is crucially important to understand how our social preferences interact and support different fairness judgments. Our research is a first step toward understanding how fairness concerns guide judgments about how, when, and in what ways we evaluate people for the way they distribute resources among others.

Acknowledgments

The authors thank Emily Gerdin, Kayla Good, and Hannah Kim as well as other members of the DIBS laboratory for critical feedback on the manuscript.

Note

1. After all participants were run in Study 1b, we ran two additional participants who were paid in line with the average choice in our two conditions from Study 1b. Thus, a real participant actually received \$2 in line with the average payoff from positive reciprocity and another real participant actually received \$1 in line with the average choice in the negative reciprocity condition. This was done to ensure that the participants' decision in Study 1b actually had consequences, which allowed us to avoid deceiving our participants in Study 1b.

References

- Abbink, K., Irlenbusch, B., & Renner, E. (2002). An experimental bribery game. *Journal of Law, Economics, and Organization*, 18, 428–454.
- Adams, J. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 2, pp. 267–299). New York: Academic Press.
- Andreoni, J. (1995). Cooperation in public-goods experiments: Kindness or confusion? The American Economic Review, 85, 891–904.
- Arasli, H., & Tumer, M. (2008). Nepotism, favoritism and cronyism: A study of their effects on job stress and job satisfaction in the banking industry of north Cyprus. *Social Behavior and Personality: An International Journal*, *36*, 1237–1250.

- Barakzai, A., & Shaw, A. (2018). Friends without benefits: When helpful friends are worse than unhelpful friends. *Evolution and Human Behavior*, *5*, 529–537.
- Barr, A., & Serra, D. (2009). The effects of externalities and framing on bribery in a petty corruption experiment. *Experimental Economics*, 12, 488–503.
- Baumard, N., André, J. B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, 36, 59–78.
- Binmore, K. (2006). Why do people cooperate? Politics, Philosophy & Economics, 5, 81-96.
- Boadi, G. E. (2000). Conflict of interest, nepotism and cronyism. In J. Pope (Ed.), *Confronting corruption: The elements of a national integrity system* (pp. 195–204). Berlin: Transparency International.
- Brick, I., Palmon, O., & Wald, J. (2006). CEO compensation, director compensation and firm performance: evidence of cronyism? *Journal of Corporate Finance*, *12*, 403–423.
- Brody, R. G., Coulter, J. M., & Mihalek, P. H. (1998). Whistle-blowing: A cross-cultural comparison of ethical perceptions of U.S. and Japanese accounting students. *American Business Review*, 16, 14–21.
- Carlsmith, K. M. (2006). The roles of retribution and utility in determining punishment. Journal of Experimental Social Psychology, 42, 437–451.
- Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 117, 817–869.
- Chiang, Y. S. (2010). Self-interested partner selection can lead to the emergence of fairness. *Evolution and Human Behavior*, 31, 265–270.
- Choshen-Hillel, S., Shaw, A., & Caruso, E. M. (2015). Waste management: How reducing partiality can promote efficient resource allocation. *Journal of Personality and Social Psychology*, 119, 210–231.
- Choshen-Hillel, S., Shaw, A., & Caruso, E. (2018). Disadvantaged but not dissatisfied: Agency interventions ameliorate negative reactions to unequal pay. *Journal of Experiment Psychology: Applied*, 24, 578–599.
- Cropanzano, R., Howes, J. C., Grandey, A. A., & Toth, P. (1997). The relationship of organizational politics and support to work behaviors, attitudes, and stress. *Journal of Organizational Behavior*, 18, 159–180.
- Damon, W. (1977). The social world of the child. San Francisco: Jossey-Bass.
- Dawes, C., Fowler, J., Johnson, T., McElreath, R., & Smirnov, O. (2007). Egalitarian motives in humans. *Nature*, 446, 794–796.
- Delton, A. W., Krasnow, M. M., Cosmides, L., & Tooby, J. (2011). Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounter. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 13335–13340.
- DeScioli, P., & Kurzban, R. (2009a). Mysteries of morality. Cognition, 112, 281-299.
- DeScioli, P., & Kurzban, R. (2009b). The alliance hypothesis for human friendship. PLoS ONE, 4, e5802.
- DeScioli, P., & Kurzban, R. (2011). The company you keep: Friendship decisions from a functional perspective. In J. I. Krueger (Ed.), *Social judgment and decision making* (pp. 205–224). New York: Psychology Press.
- DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin*, 139, 477–496.
- DeScioli, P., Massenkoff, M., Shaw, A., Petersen, M. B., & Kurzban, R. (2014). Equity or equality? Moral judgments follow the money. *Proceedings of the Royal Society B: Biological Sciences*, 281(1797), 20142112–20142112.
- Deutsch, M. (1975). Equity, equality, and need: What determines which value will be used as the basis of distributive justice. *Journal of Social Issues*, 31, 137–149.
- Dungan, J., Waytz, A., & Young, L. (2014). Corruption in the context of moral trade-offs. Journal of Interdisciplinary Economics, 26, 97–118.
- Elster, J. (2006). Fairness and norms. Social Research, 73, 365-376.
- Everett, J. A., Faber, N. S., Savulescu, J., & Crockett, M. J. (2018). The costs of being consequentialist: Social inference from instrumental harm and impartial beneficence. *Journal of Experimental Social Psychology*, 79, 200–216.

- Falk, A., Fehr, E., & Fischbacher, U. (2008). Testing theories of fairness—Intentions matter. *Games and Economic Behavior*, 62, 287–303.
- Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. Games and Economic Behavior, 54, 293-315.
- Fehr, E., & Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. Journal of Economic Perspectives, 14, 159–181.
- Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition and cooperation. *Quarterly Journal of Economics*, 114, 817–868.
- Festinger, L. (1954). A theory of social comparison processes. Human Relations, 7, 117-140.
- Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review*, *99*, 689–723.
- Fiske, A. P., & Rai, T. S. (2014). Virtuous violence: Hurting and killing to create, sustain, end, and honor social relationships. Cambridge: Cambridge University Press.
- Gordon-Hecker, T. D., Rosensaft, D., Pittarello, A., Shalvi, S., & Bereby-Meyer, Y. (2017). Not taking responsibility: Equity trumps efficiency in allocation decisions. *Journal of Experimental Psychology: General*, 146, 771–775.
- Gray, K., Ward, A. F., & Norton, M. I. (2014). Paying it forward: Generalized reciprocity and the limits of generosity. *Journal of Experimental Psychology: General*, 143, 247–254.
- Gurven, M. (2003). Reciprocal altruism and food sharing decisions among Hiwi and Ache hunter-gatherers. *Behavioral Ecology and Sociobiology*, *56*, 366–380.
- Gurven, M. (2006). The evolution of contingent cooperation. Current Anthropology, 47, 185–192.
- Gurven, M., & Winking, J. (2008). Collective action in action: Prosocial behavior in and out of the laboratory. American Anthropologist, 110, 179–190.
- Haidt, J., & Joseph, C. (2008). The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind* (Vol. 3, pp. 367–392). New York: Oxford University Press.
- Harstad, B., & Svensson, J. (2011). Bribes, lobbying, and development. *American Political Science Review*, 105, 46–63.
- Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C, Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker, J. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science*, 327, 1480–1484.
- Hill, K., & Kaplan, H. (1993). On why male foragers hunt and share food. *Current Anthropology*, 34, 701–710.
- Hook, J., & Cook, T. D. (1979). Equity theory and the cognitive ability of children. *Psychological Bulletin*, 86, 429–445.
- Hsee, C. K. (1996). Elastic justification: How unjustifiable factors influence judgments. Organizational Behavior and Human Decision Processes, 66, 122–129.
- Hughes, J. S. (2017). In a moral dilemma, choose the one you love: Impartial actors are seen as less moral than partial ones. *British Journal of Social Psychology*, *56*, 561–577.
- Keysar, B., Converse, B. A., Wang, J., & Epley, N. (2008). Reciprocity is not give and take: Asymmetric reciprocity to positive and negative acts. *Psychological Science*, 19, 1280–1286.
- Khatri, N., & Tsang, E. W. (2003). Antecedents and consequences of cronyism in organizations. *Journal of Business Ethics*, 43, 289–303.
- Kiyonari, T., & Barclay, P. (2008). Cooperation in social dilemmas: Free riding may be thwarted by second-order reward rather than by punishment. *Journal of Personality and Social Psychology*, 95, 826–842.
- Kleiman-Weiner, M., Shaw, A., & Tenenbaum, J. B. (2017). Constructing social preferences from anticipated judgments: When impartial inequity is fair and why? In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 676– 681). Austin, TX: Cognitive Science Society.

- Kleinfeld, N. R. (2014). A bridge to scandal: Behind the fort lee ruse. *The New York Times*, January 12. Retrieved from https://www.nytimes.com/2014/01/13/nyregion/a-bridge-to-scandal-behind-the-fort-lee-ruse. html?_r=0. Accessed February 07, 2018.
- Kunda, Z. (1990). The case for motivated reasoning. Psychological Bulletin, 108, 480-498.
- Litman, L., Robinson, J., & Abberbock, T. (2016). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, 49, 433–442.
- Ma, L. K., Tunney, R. J., & Ferguson, E. (2014). Gratefully received, gratefully repaid: The role of perceived fairness in cooperative interactions. *PLoS ONE*, 9, e114976.
- MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, 58, 593–614.
- Messick, D. M. (1995). Equality, fairness, and social conflict. Social Justice Research, 8, 153-173.
- Messick, D. M., & Sentis, K. P. (1979). Fairness and preference. Journal of Experimental Social Psychology, 15, 418–434.
- Miller, J. G., & Bersoff, D. M. (1992). Culture and moral judgment: How are conflicts between justice and interpersonal responsibilities resolved? *Journal of Personality and Social Psychology*, 62, 541–554.
- Murnighan, J. K. (1978). Models of coalition behavior: Game theoretic, social psychological, and political perspectives. *Psychological Bulletin*, 85, 1130–1153.
- Niemi, L., Wasserman, E., & Young, L. (2018). The behavioral and neural signatures of distinct conceptions of fairness. Social Neuroscience, 13, 399–415.
- Niemi, L., & Young, L. (2017). Who sees what as fair? Mapping individual differences in valuation of reciprocity, charity, and impartiality. *Social Justice Research*, 30, 438–449.
- Offerman, T. (2002). Hurting hurts more than helping helps. European Economic Review, 46, 1423–1437.
- Pinker, S. (2008). The Moral Instinct. *The New York Times*, January 13. Retrieved from http://www.nytimes.c om/2008/01/13/magazine/13Psychology-t.html. Accessed April 12, 2018.
- Poon, J. M. (2004). Effects of performance appraisal politics on job satisfaction and turnover intention. *Personnel Review*, 33, 322–334.
- Rasinski, K. A. (1987). What's fair is fair—Or is it? Value differences underlying public views about social justice. Society of London B: Biological Sciences, 281, 20142112.
- Ray, D. (2007). A game-theoretic perspective on coalition formation. Oxford, UK: Oxford University Press.
- Reinsch, N. L., & Gardner, J. A. (2014). Do communication abilities affect promotion decisions? Some data from the C-Suite. *Journal of Business and Technical Communication*, 28, 31–57.
- Rubin, J. Z., Pruitt, D. G., & Kim, S. H. (1994). Social conflict: Escalation, stalemate, and settlement. New York: McGraw-Hill.
- Scholl, B. J. (2017). Reliability in psychology: Means versus ends. APS Observer, 30, 38-39.
- Schroeder, D. A., Steel, J. E., Woodell, A. J., & Bembenek, A. F. (2003). Justice within social dilemmas. Personality and Social Psychology Review, 7, 374–387.
- Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-serving justifications: Doing wrong and feeling moral. *Current Directions in Psychological Science*, 24, 125–130.
- Shaw, A. (2013). Beyond "To share or not to share": The impartiality account of fairness. Current Directions in Psychological Science, 22, 413–417.
- Shaw, A. (2016). Fairness: What it isn't, what it is, and what it might be for. In D. C. Geary & D. B. Berch (Eds.), Evolutionary perspectives on child development and education (pp. 193–214). Basel, Switzerland: Springer International Publishing.
- Shaw, A., & Choshen-Hillel, S. (2017). It's not fair: Folk intuitions about disadvantageous and advantageous inequity aversion. Judgment and Decision Making, 12, 208–223.
- Shaw, A., Choshen-Hillel, S., & Caruso, E. (2016). The development of inequity aversion: Understanding when (and why) people give others the bigger piece of the pie. *Psychological Science*, *27*, 1352–1359.
- Shaw, A., Choshen-Hillel, S., & Caruso, E. M. (2018). Being biased against friends to appear unbiased. Journal of Experimental Social Psychology, 78, 104–115.

- Shaw, A., DeScioli, P., Barakzai, A., & Kurzban, R. (2017). Whoever is not with me is against me: The costs of neutrality among friends. *Journal of Experimental Social Psychology*, 71, 96–104.
- Shaw, A., DeScioli, P., & Olson, K. R. (2012). Fairness versus favoritism in children. Evolution and Human Behavior, 33, 736–745.
- Shaw, A., & Knobe, J. (2013). Not all mutualism is fair and not all fairness is mutualistic. *Behavior and Brain Sciences*, *36*, 100–101.
- Shaw, A., & Olson, K. R. (2012). Children discard a resource to avoid inequity. Journal of Experimental Psychology: General, 141, 382–395.
- Shaw, A., & Olson, K. R. (2014). Fairness as an aversion to partiality: The development of procedural justice. *Journal of Experimental Child Psychology*, 119, 40–53.
- Snyder, G. H. (1984). The security dilemma in alliance politics. World Politics, 36, 461-495.
- Sznycer, D., Seal, M. F. L., Sell, A., Lim, J., Porat, R., Shalvi, S., Halperin, E., Cosmides, L., & Tooby, J. (2017). Support for redistribution is shaped by compassion, envy, and self-interest, but not a taste for fairness. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 8420– 8425.
- Thibaut, J. W., & Walker, L. (1975). *Procedural justice: A psychological analysis*. Hillsdale, NJ: Erlbaum Associates.
- Tyler, T. R. (2000). Social justice. International Journal of Psychology, 35, 117-125.
- Van Beest, I., Van Dijk, E., De Dreu, C. K., & Wilke, H. A. (2005). Do-no-harm in coalition formation: Why losses inhibit exclusion and promote fairness cognitions. *Journal of Experimental Social Psychology*, 41, 609–617.
- Van Beest, I., Wilke, H., & Van Dijk, E. (2003). The excluded player in coalition formation. *Personality and Social Psychology Bulletin*, 29, 237–247.
- Waytz, A., Dungan, J., & Young, L. (2013). The whistleblower's dilemma and the fairness-loyalty tradeoff. Journal of Experimental Social Psychology, 49, 1027–1033.
- Weisel, O., & Shalvi, S. (2015). The collaborative roots of corruption. Proceedings of the National Academy of Sciences of the United States of America, 112, 10651–10656.
- Yamagishi, T., Jin, N., & Miller, A. S. (1998). In-group bias and culture of collectivism. Asian Journal of Social Psychology, 1, 315–328.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Appendix S1. Supplemental studies.