Northwestern | INSTITUTE FOR POLICY RESEARCH

IPR Working Paper Series

WP-23-01

Collaboration Induces Debt-Motivated Altruism

Mary McGrath
Northwestern University and IPR

Version: January 17, 2023

DRAFT

Please do not quote or distribute without permission.

Abstract

Collaboration with others—even a minimal instance—increases willingness to sacrifice on their behalf. What is the mechanism underlying this relationship? An increased willingness to sacrifice could arise from a general desire to improve the other's wellbeing, from a norm-bound sense of debt owed to one's collaborator, or (even after controlling for other egoistic concerns) from an aim towards a "warm glow" feeling from making the sacrifice. Understanding the mechanism at work is not simply a matter of theoretical interest, but of crucial importance in understanding broader implications of the collaboration effect and how it alters our relationships with others. This paper presents results from four randomized experiments investigating this mechanism. Rather than exhibiting egoistic concerns, a desire to strictly increase the collaborator's wellbeing, or a general aversion to inequality, people behave as if collaboration creates an obligation of debt owed to the collaborator. Taken together, the evidence from these experiments suggests that collaboration produces a bounded form of altruism, focused on what is due but not beyond.

Data and code for producing the analyses and figures are accessible in Northwestern University's Arch Data Repository: https://arch.library.northwestern.edu/collections/bz60cw536. This research was supported by the Center for the Study of American Politics at Yale University, and the Institute for Policy Research at Northwestern University.

Collaboration induces debt-motivated altruism

I. Introduction

From the start of the COVID-19 pandemic, public officials and private individuals appealed to the mantra, "We're in this together," as a phrase intended not only to provide solace, but also to promote endurance—the resolve to bear up as governments asked the public to make sacrifices on a massive scale. The ubiquity of this appeal to collective experience suggests many found the association natural: being part of a shared endeavor entails a responsibility to make necessary sacrifices for the others involved. At the same time, the glaring disparities in the pandemic's health and economic effects made painfully clear the all-important ways in which people were *not* in it together. To the extent that recourse to this collective appeal resonated in conditions of such tenuous "togetherness" suggests a motivating force that is reflexive more than rational.

What underlies this association between collective endeavor and willingness to sacrifice? A *collaborator principle*—by which working in common cause with someone engenders an increased willingness to bear costs on their behalf—appears to be a distinct phenomenon, unique to humans (Hamann et al., 2011; McGrath & Gerber, 2019). *Why* does collaboration have this effect? Different underlying mechanisms causing a behavior can carry vastly different downstream implications (see e.g., DellaVigna et al., 2012; Malmendier et al., 2014)—and public appeals invoking this effect to promote prosocial behavior make clear the real-world importance of understanding the nature of this collaboration effect.

This study presents an investigation of the mechanism behind the collaboration effect. I hypothesize that the collaboration effect operates by creating a sense of indebtedness to one's collaborator, generating a bounded form of altruism. In four experiments designed to shed light on the mechanism, I find behavioral evidence supporting this hypothesis.

II. Background – prosocial behavior

Scope of the study

Tinbergen, 1963 identified four main questions that orient the study of behavior—those of mechanism (how does it work?); adaptive function (what does it do?); ontogeny (how does it develop in the individual?); and phylogeny (what is its evolutionary history?). These questions are distinct and complementary—though interrelated, an answer to any one of these questions cannot be considered as also answering one of the others (Bateson and Laland, 2013). The orienting structure provided by these questions is perhaps especially useful in the study of prosocial behavior, which spans disparate fields and a range of objectives.

The aim of this study is to provide empirical evidence on the psychological *mechanism* underlying the prosocial effect of collaboration. Collaboration increases willingness to sacrifice (Hamann et al., 2011; McGrath and Gerber, 2019)—how does this work? While an integration of findings from all four areas of focus should be an ultimate aim in the study of behavior (Tinbergen, 1963), for present purposes the evolutionary emergence of this phenomenon, its adaptive purpose, and how it develops over the lifespan are set aside. Specifying this focus is important here not only to clarify the scope of the current investigation, but also because the definition of basic terms (e.g., "altruism") can differ depending on which question is at hand (see Hawley, 2014).

The remainder of this Background section contextualizes the collaboration effect within the broader study of prosocial behavior, defining key terms in this investigation of the mechanism. The subsequent section (III. Theory) presents the theoretical predictions that lay the foundation for the experimental tests (IV. Methods; V. Results) that follow.

Varieties of prosocial behavior

Prosocial behavior is a voluntary action with knowing benefit to another. An involuntary action—a common example is paying taxes—does not constitute prosocial

behavior by this definition. Likewise, a behavior is not prosocial if the benefit to another is unknowing—for example, throwing a half-eaten sandwich in the garbage does not become prosocial if a hungry person finds and eats it.[1]

Prosocial behavior can be characterized by the intersection of costs and benefits to the actor, distinguishing four varieties: *cooperation, collaboration, sacrifice*, and *decency*. With *cooperation*, "people pay costs to benefit others" (Bear and Rand, 2016), but also partake in some shared benefit resulting from the action. In contrast, *collaboration* itself imposes no direct costs on the actor, but still confers to that actor some shared benefit resulting from the action (McGrath and Gerber, 2019; Tomasello and Vaish, 2013). *Sacrifice*, on the other hand, is a prosocial behavior in which the actor incurs costs but draws *no* direct benefit (see, e.g., Impett et al., 2005). A fourth type can be referred to as *decency*: a prosocial behavior from which the actor gains no direct benefit, but also bears no costs (e.g., calling out to a passerby unaware they dropped a glove; giving a commuter your half-spent subway card at the end of your stay in a city).

Cooperation is the most commonly studied form of prosociality. Widespread in human behavior, cooperation presents a puzzle interesting to researchers from evolutionary theory to economics: why bear costs—especially when, as in many circumstances, immediate gain would be greater from non-cooperation? Studies of cooperation often focus on understanding its adaptive function or phylogeny.

The present study does not look at cooperation, but instead examines the relationship between two other varieties of prosocial behavior: collaboration and sacrifice. Sacrifice, as indicated above, is a costly prosocial action that provides no direct benefit to the actor (see also Van Lange et al., 1997). *Collaboration* is working together toward a

---

[1] I use the criterion of *knowing* benefit to another (i.e., the actor is aware of the benefit provided) in place of a more commonly used criterion of *intent* (e.g., Hawley, 2014) because it provides a more clear-cut, if broader, standard. *Voluntary* remains difficult to delineate: giving your watch to a mugger obviously does not constitute prosocial behavior; volunteering at a food bank to fulfill a mandatory course requirement is more ambiguous—but for present purposes would not meet the *voluntary* criterion.

shared goal. This shared goal further differentiates collaboration from cooperation; with cooperation, actors have separate but identical goals (as in a prisoner's dilemma, where each wants to minimize their own time served) or separate and distinct goals (e.g., a witness cooperating with federal investigators). Whereas collaboration entails a process of joint effort ("working together") towards the shared goal, cooperation is characterized by a process of exchange, or *quid pro quo*.

Altruism

Prosocial behavior is motive-neutral: a voluntary action with knowing benefit to another is prosocial regardless of the actor's motivation. In other words, a behavior can be prosocial even if it is driven by a purely self-centered motivation (see Hawley, 2014).

A motivation is a goal-directed, conscious or unconscious internal psychological force (Bargh and Morsella, 2008; Batson and Shaw, 1991; Hawley, 2014).[2] An *other-regarding motivation* is a motivation wherein the goal in question centers around an "other"—in contrast to an egoistic motivation, wherein the goal centers around the self. More specifically, a motivation is *altruistic* if the ultimate goal is improving another's welfare, and *egoistic* if the ultimate goal is improving one's own welfare (Batson and Shaw, 1991). *Altruism*, in this context, is prosocial behavior driven by an altruistic motivation (Hawley, 2014).

Motivations that can lead toward an ultimate goal of improving another's welfare include empathy (Batson et al., 2002; FeldmanHall et al., 2015), affinity (Moreland and Zajonc, 1982), gratitude (Bartlett and DeSteno, 2006; Tsang, 2006), and obligation (Coleman, 1988; Dahl and Paulus, 2019; Greenberg, 1980; Tomasello, 2020). These motivations can produce altruism that is bounded (limited by the contours of a norm—e.g., as when motivated by obligation) or unbounded (e.g., as when motivated by affinity), and

---

[2] Reeve, 2018 categorizes three specific types of "internal psychological force" that serve as motivations: needs, cognitions, and emotions.

contingent (tied directly to the value of an antecedent—e.g., as when motivated by gratitude) or non-contingent (e.g., as when motivated by empathy).[3]

Reciprocity

Reciprocity is the behavior arising from the decision-rule, "respond in kind," or *reciprocate* (see Levine, 1998; Rabin, 1993). This decision-rule can—but does not necessarily—produce prosocial behavior. For example, a decision-rule of destructive reciprocity ("repaying unkindness with unkindness"; Sobel, 2005, p.397) could prescribe purely vindictive retribution, inflicting harm without benefiting any other person. Because there is no benefit to any other person, this would not constitute prosocial behavior. The same decision-rule could produce prosocial behavior if employed for norm-enforcing punishment that serves to benefit others (i.e., "strong reciprocity"; see Bowles and Gintis, 2004; Fong et al., 2006). When reciprocity provides knowing benefit to another, it constitutes contingent prosociality.[4] When the ultimate goal of that reciprocity is to improve the other's welfare, it constitutes contingent altruism.

Figure 1 uses the example of reciprocity to map the connections between these concepts. This figure illustrates a number of key points: prosocial reciprocity can be driven by an egoistic motivation or an other-regarding motivation, including the examples of altruistic motivations discussed above. For example, reciprocity with an ultimate goal of improving the recipient's welfare could be motivated by gratitude (producing a contingent and unbounded form of altruism) or by obligation (producing a contingent and bounded form of altruism).

---

[3] The more generic term "conditional altruism" does not distinguish between conditions placed on the altruism (i.e., bounded altruism) and conditions placed on the recipient (i.e., contingent altruism).

[4] In other words, prosocial reciprocity is a form of contingent prosociality. Note that contingent prosociality is a broader category than just prosocial reciprocity, as the contingency could hinge on a decision-rule other than "respond in kind" (i.e., depend on a different antecedent). For example, parochial altruism would also constitute a form of contingent prosociality, contingent on in-group/out-group status of the recipient.

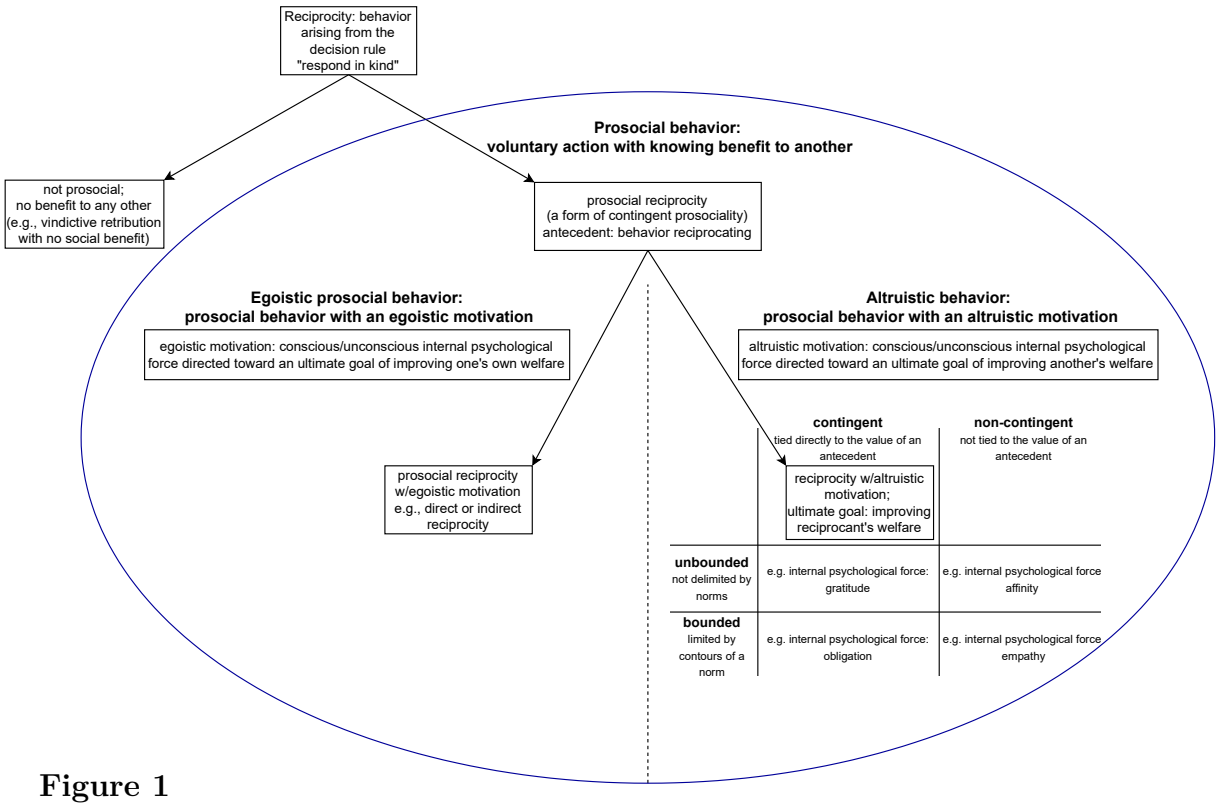

Figure 1

Mapping of reciprocal behavior into prosociality. When the ultimate goal of reciprocal behavior is to improve another's welfare, that behavior constitutes contingent altruism. Contingent altruism can be bounded (circumscribed by the contours of a norm) or unbounded (not delimited by norms).

Malmendier et al., 2014 note the importance of distinguishing underlying motives behind reciprocal behavior: prosocial reciprocity may arise from a *desire* to improve the other person's welfare, or because those reciprocating "feel obliged or pressured" to improve the other's welfare (p. 863).[5] Greenberg, 1980 divides the motivational bases for

[5] Malmendier et al., 2014, writing from an economic perspective, focus on the distinction between internal and external motivations. However, because from a psychological perspective all motivation is internal (and always influenced by external factors) (see Reeve, 2018), a more fitting distinction is between motivations that represent a desire or pull ("attractive," to use Greenberg's (1980) term) and those that present a pressure or push ("normative," in Greenberg's terminology). This distinction fits with Malmender et al.'s (2014) findings. Integrating this with the framework set out here, pull/attractive

reciprocity into three categories: future gain (i.e., egoistic); attraction to the benefactor (and a concomitant desire to express regard, interact with, and generally tend to that benefactor's wellbeing); and indebtedness to the benefactor. Greenberg writes that indebtedness—a state of obligation to repay another—can produce a form of normative reciprocity, not directed toward future benefits or dependent on attraction, but instead "motivated by the feeling of obligation" (Greenberg, 1980, p. 15). This reciprocity of obligation differs from other forms of reciprocity, both theoretically and empirically (see Tomasello, 2020; Warneken and Tomasello, 2013). Direct and indirect reciprocity, for example, are motivated by anticipation of benefits to the self (Axelrod & Hamilton, 1981; Nowak & Sigmund, 2005) (see Fig. 1). And the normative nature of reciprocity resulting from indebtedness differs from the role of norms in strong reciprocity, wherein costly aid or punishment is dependent upon a potential recipient's adherence to a norm (Fong et al., 2006): with indebtedness, the force of the norm falls on the donor (i.e., the debtor), not the potential recipient.

A reciprocity of indebtedness also differs from the prosocial reciprocity motivated by gratitude (Watkins et al., 2006; see also Tomasello, 2020), which would fall into Greenberg's category of attraction (desire to express regard for, interact with, or increase the wellbeing of a benefactor). Whereas gratitude is associated with positive affect and is directed toward developing social relationships (e.g., a desire for closeness with the helper, Salovey et al., 1991), indebtedness is associated with negative or mixed affect, and an obligation to compensate another with the goal of restoring equity (Peng et al., 2018). Because indebtedness involves a focus on repaying the cost of a debt, even when the goal is improving another's welfare (up to the point of recompense) a reciprocity of indebtedness may be expected to aim for fulfillment of the obligation, but not beyond (Tsang, 2006).

Tomasello (2020) writes that a sense of obligation, which is unique to humans, arises

---

motivations generate unbounded altruism (not delimited by the contours of a norm) while push/normative motivations generate bounded altruism (delimited by the norm in question).

from our ability to enter into collaborations—a process of shared intentionality deeply intertwined with the development of both culture and morality. Humans hold the ability to create a collective, self-regulating agent—*we*—that carries with it expectations of how the individuals within that collective will treat each other. That *we* can incorporate not only direct collaborators, but anyone who adheres to the collective commitments—i.e., a cultural group. Those collective expectations of behavior then take on an "objective" weight, as moral obligations. These obligations extend only as far as the collective *we*: "the sense of obligation (in contrast to the sense of sympathy) only operates within one's moral community" (Tomasello, 2020).

Tomasello's work suggests that a sense of obligation arising from collaboration may lie at the root of human morality, and help to define the boundaries of what we owe to whom. This deep-seated relationship could account for our seemingly reflexive response to collaboration—the apparently natural association between common endeavor and willingness to sacrifice.

III. Theory – how do collaboration effects operate?

Figure 1 illustrates key takeaways from the background set out above: prosocial reciprocity can be driven by an egoistic motivation (Axelrod & Hamilton, 1981; Malmendier et al., 2014; Nowak & Sigmund, 2005) or an other-regarding motivation (Bowles & Gintis, 2013; Greenberg, 1980; Levine, 1998; Sobel, 2005; Sugden, 1984), and different forms of other-regarding motivations are possible as well—e.g., a desire to improve another's welfare vs. an obligation to repay a debt. These different motivations carry different behavioral and psychological implications (see DellaVigna et al., 2012; Impett et al., 2005).

Successful collaboration, even a minimal instance pared to its essence, induces a form of prosocial behavior: it increases willingness to sacrifice for one's partner (Hamann et al., 2011; McGrath & Gerber, 2019). The aim of the present investigation is to understand the nature of this effect. *Why* does collaboration increase willingness to

sacrifice?

The literature above suggests three possibilities. Collaboration may increase prosocial behavior toward the collaborator through a sense of gratitude or affinity arising from the act of collaborating—a desire to improve the other's welfare that produces an unbounded altruism, similar to that produced by bonds of kinship (Becker, 1974). Alternatively, collaboration may create a sense of indebtedness to the collaborator, such that the increased prosociality is driven by obligation—the norm-bound reciprocity described by Greenberg, 1980. Yet another possibility is that the prosocial behavior produced by collaboration is not altruistic at all, but egoistic—centered around one's own welfare rather than the welfare of the other person. Even without externally-facing considerations like reputation, internally-facing egoism—e.g.,"warm-glow" giving (Andreoni, 1990)—could produce prosocial behavior.

I hypothesize that collaboration increases willingness to sacrifice by creating a sense of indebtedness to the collaborator, producing a bounded and contingent altruism motivated by obligation.

Theoretical models of these three potential mechanisms predict divergent behavior in response to manipulation of how cost borne by the actor confers benefit to the recipient. These differentiating predictions provide a behavioral test that can shed light on the mechanism at work.

Predictions for egoism, unbounded altruism, and bounded altruism

Say that a potential donor has an endowment E; that X is the amount of E she keeps for herself; and that x is the amount of E she gives to a recipient. With a model of unbounded altruism (see, e.g., Konow's model of "pure altruism"; Konow, 2010), a donor's utility is a monotonic function of her own allocation and the utility of another, the recipient. The donor's utility can be written:

$$u(X) + v(h(x)), \tag{1}$$

where $u(X)$ represents the donor's material utility from X, $h(x)$ represents the recipient's utility from the gift x, and $v(h(x))$ represents the utility the donor receives from the recipient's utility from x. For simplicity, we can equivalently write $f(x) = v(h(x))$. The functions $u()$ and $f()$ are assumed to be twice continuously differentiable and to provide positive but diminishing marginal utility. It is also assumed that $-(x^* f'')/f' < 1$, meaning that such altruists are "magnanimous," i.e., that $f'()$ is inelastic. (See *SI Appendix 2.*)

With warm-glow giving (which is purely egoistic, Andreoni, 1990), the donor's utility is a function of her own allocation and the size of her donation to the recipient:

$$u(X) + g(x), \tag{2}$$

where $u(X)$ is again the donor's material utility, and $g(x)$ is the "warm glow" gain she receives from giving a donation of amount x (Andreoni, 1990). In contrast to unbounded altruism, the recipient's utility, $h(x)$, does not appear in a warm-glow donor's utility function; the donor's utility derives strictly from the amount the donor gives up, with no attention given to the utility gained by the recipient. It is assumed that $g()$ is twice continuously differentiable and provides positive but diminishing marginal utility.

Bounded altruism introduces a qualifying norm, ϕ, representing the normative amount for the recipient to receive (see Konow, 2010). This norm is incorporated into the donor's utility function such that deviation of the donor's gift from that norm provides disutility to the donor:

$$u(X) - f(x - \phi), \tag{3}$$

where it is now assumed that $f'(x - \phi)(x - \phi) > 0$ and $f''() > 0$. This implies that $-f()$ takes its maximum where the donor gives the normative amount. Here, the motivating norm is repayment of a debt. ϕ represents the "right" amount to give according to the motivating norm (Konow, 2010). In this case, the right amount is the debt owed to the collaborator. Importantly, the donor's objective is not to generally improve the recipient's well-being, as it is with unbounded altruism, but to meet (and not exceed) some amount due.

Given these functional forms, a "matching grant"—through which a given level of cost borne by the actor confers greater benefit to the recipient—provides a useful tool for distinguishing between models. An unbounded or warm-glow gift is non-decreasing with a multiplier on the gift in the form of a matching grant.[6] With warm-glow giving, a matching grant should have no effect: because only the size of the gift, not the recipient's utility from the gift, matters to the warm-glow donor, the matching grant does not appear in the utility function.[7] With unbounded altruism, the matching grant does appear in the utility function: if $\kappa > 1$ is the matching grant multiplier, utility is $u(X) + f(\kappa x)$ for an unbounded altruist. Here, giving should *increase* with the size of a matching grant multiplier. The same would be true for an "impure altruist," i.e., a donor who receives utility from her own allocation, from the recipient's utility, and from the size of her donation. (*SI Appendix 2* demonstrates these results.)

In contrast, a matching grant can *decrease* the amount given when a donor is debt-motivated (bounded altruism). Utility for a debt-motivated donor under a matching grant is $u(X) - f(\kappa x - \phi)$. Exceeding repayment of the debt detracts from the debt-motivated donor's utility both in terms of reduced material utility (X is smaller than necessary) and through a penalty for exceeding the norm (deviation of κx from ϕ).

With bounded altruism, a sufficient condition for a matching grant to decrease donations occurs when giving in the standard (non-matching) condition is greater than ϕ/κ

---

[6] Note that this differs from predictions of crowding out with an unrestricted external subsidy to the recipient or with a tax-funded grant to the recipient (see Konow, 2010). *SI Appendix 2* provides a proof. Empirical evidence from the literature on charitable giving supports a non-negative effect of lowering the price of donation through a matching grant (see Karlan and List, 2007, which also reports heterogeneity that would correspond with warm-glow donation in blue states and output-focused donation in red states).

[7] If the warm-glow donor were to conceive of the matching grant as increasing the size of her gift, so that the matching grant increases her "warm glow" as opposed to only increasing the amount received by the recipient, then the multiplier appears in $g(\kappa x)$ and warm-glow giving would increase with the matching grant, as is the case with unbounded and impure altruism.

(see *SI Appendix 3*). When this is true, unless the matched donor gives less than the standard donor, the matching grant recipient will receive an amount that exceeds the norm—in this case, the debtee will receive an amount that exceeds repayment of the debt. The donor's utility decreases because she has given up more than was required. In the present experiments, $\kappa = 2$ and all participants earned a \$0.50 award from which they could donate. Earlier findings suggest that under the design employed, the collaboration-induced norm would be to give half of the bonus payment, i.e., here, $\phi = \$0.25$ (Konow, 2010; McGrath & Gerber, 2019). If donations in the standard condition exceed $\phi/\kappa = \$0.125$, a matching grant is able to decrease giving by debt-motivated respondents.

Experiments 1 & 2 provide a test of these predictions.

Inequity aversion and inequality aversion

Inequity aversion is a hypothesis proposed by the economists Fehr and Schmidt, 1999 (and contemporaneously Bolton and Ockenfels, 2000) as an alternative to the assumption of purely self-interested preferences. This hypothesis holds that some people (under some circumstances, Fershtman et al., 2012) will move toward more "equitable" outcomes, even at a cost to themselves (Fehr and Schmidt, 1999). In other words, the hypothesis of inequity aversion posits that at least some people act not only to maximize their own material outcomes, but (also or instead) to achieve what they perceive as fair. A preference for "what is fair" is accommodated in the models above via ϕ, perception of the normatively "right" amount in the model of bounded altruism. Inequity aversion itself is captured by $f'(x-\phi)(x-\phi) > 0$ and $f''() > 0$.

Note that inequity aversion is not a motivation,[8] but a behavior: movement toward

[8] Some scholars refer to inequity aversion as a motivation, and an argument can be made for the existence of "principlism" as a motivating force wherein the goal centers on adherence to an abstract principle (e.g., equity). However, see Batson et al., 2002, who note that while the upholding of a principle is theoretically possible as an ultimate goal, the balance of evidence has indicated that apparent principlism is more often in service to egoistic, altruistic, or collectivist goals. (For example, few people would be committed to

more "equitable" outcomes. Indeed, inequity aversion comprises at least two wholly distinct behavioral phenomena—Advantageous Inequity aversion and Disadvantageous Inequity aversion—that exhibit separate ontogenies and are supported by different psychological mechanisms (see Blake et al., 2015; Corbit et al., 2017).[9]

Based on $f'(x-\phi)(x-\phi) > 0$ and $f''() > 0$ in the model of bounded altruism, and setting $\phi = \$0.25$ based on prior empirical evidence, testing for the presence or absence of "inequity averse" behavior in a specified context is useful for distinguishing between models. However, while evidence of this behavior under predicted conditions can be used to distinguish between the models set out above, the presence of this behavior does not imply that the effect of collaboration can be wholly characterized as inequality aversion.[10]

In short, the usefulness of inequity averse behavior as a diagnostic does not imply that inequality aversion drives the collaboration effect (it is an outcome, not a motivation), or that the effect of collaboration is synonymous with inequality aversion. On the contrary, my hypothesis—that collaboration creates a sense of something owed to the collaborator—implies that the collaboration effect is neither a product of nor an incarnation of inequality aversion, and that collaboration should *not* increase inequality aversion generally, and is likely to decrease it in some circumstances. For example, if the increased willingness to sacrifice arises from a sense of indebtedness to one's collaborator, (i) this should not increase willingness to take from one's collaborator in order to equalize

---

equity even at the expense of their own and everyone else's welfare.) More to the point for present purposes, while principlism as a motivating force is theoretically possible, inequity aversion as a principlistic motivation is ruled out by Experiments 3 & 4.

[9] Further broadening the scope of the term, "equity" is generally undefined except in the specific subset of cases where it is assumed to mean equality. In other words, because equity—"what's fair"—is defined by the perception of the actor involved (Fehr and Schmidt, 1999), essentially any behavior could in theory be defended as a pursuit of equity.

[10] To suggest this would be akin to saying: A is evidence of B, therefore A is the same thing as B. To put it another way, smoke (the diagnostic evidence) is not the same thing as fire (the subject of interest).

outcomes (and may decrease willingness to take); and (ii) this increased willingness to sacrifice should be apparent in the absence of unequal outcomes (i.e., should manifest even when there is no inequality to avert).

Experiments 3 & 4 provide tests of these predictions.

IV. Methods

The theoretical predictions about warm glow, unbounded, and bounded altruism allow for the introduction of a matching grant as a test for distinguishing motives. When sharing behavior reflects warm-glow giving, unbounded altruism, or impure altruism, introducing a matching grant should not decrease the amount given. When sharing is motivated by repayment of a debt—an instance of bounded altruism—a matching grant can be expected to decrease the amount given.

To test whether the willingness to sacrifice earned resources after collaborating appears more reflective of a general desire to improve the other's welfare (unbounded altruism), a self-focused desire to feel generous (warm-glow egoism), or an obligation to repay a debt (bounded altruism), I conduct and replicate a collaboration experiment with a matching grant (Experiments 1 & 2). In two additional experiments, I show that collaboration does not generally increase inequality aversion, indicating that the prosocial effect of collaboration cannot be explained as simply a desire to equalize outcomes (Experiments 3 & 4).

These experimental designs intentionally employ a minimal, stripped-to-the-essence instance of collaboration, operationalized as a shared goal and interdependence of outcomes. Collaborating partners work toward a joint goal with interdependent outcomes: each collaborating partner's success depends on the other's work and success. Separately-working partners perform the same task as collaborating partners and as each other, but have independent goals and independent outcomes: each separately-working partner's outcome depends solely on their own work. In both conditions, partners are not

given any opportunity to meet or interact.

Limiting treatment to a form of collaboration pared down to its minimal elements in this way serves to isolate the treatment effect estimates from confounding influences that could arise from more complexly defined collaborative interactions (e.g., exclusion restriction violations arising from differences in the nature of interpersonal interaction between collaboratively-working and separately-working partners).

Prior work has shown that even such minimal instances of collaboration increase willingness to sacrifice (McGrath and Gerber, 2019). The aim of the present experiments is to investigate the mechanism underlying this fundamental effect of collaboration. Certainly, collaborations in real-world settings are vastly more complex—people decide whether and when to collaborate, choose with whom to collaborate, have positive and negative interpersonal reactions to their collaborators, and may anticipate opportunities to interact with their collaborators in the future. These and countless other factors undoubtedly complicate the effects of collaboration. However, a minimalist paradigm is necessary in trying to understand the elemental relationship underlying complex real-world phenomena. Why does collaboration, in this essential form, increase willingness to sacrifice? What can we discern about the mechanism driving this effect?

Procedure

In all four experiments, participants complete a data entry task for pay, randomly assigned to either a collaboratively-working or a separately-working condition. For those in the collaboratively-working condition, the payment schedule depends on the partner's work as well as one's own work. In the separately-working condition, the payment schedule does not depend on the partner's work, only on one's own work. The design controls across conditions for considerations of work effort, inequality, reputation, and anticipation of personal gain (see *SI Appendix 1: Extended Methods*).

In Experiments 1 and 2, after the work task participants are presented with an

opportunity to give some of their earnings to their partner, who, in a lottery draw, received less than the participant. For half the participants in each condition, this opportunity is presented as a "matching grant," in which the partner would receive double the amount given up by the participant.

In Experiments 3 and 4, participants are either given an opportunity to give some of their earnings to a partner who has received the same amount as the participant in a lottery draw (Experiment 3 - baseline equality), or are given the opportunity to take some surplus from a partner who has received more than the participant in the lottery draw (Experiment 4 - disadvantageous inequality).

Hypotheses

H1: The matching grant will have no effect in a separately-working condition.

In the separately-working condition, two people work on the same task in parallel, with independent outcomes. There is no collaboration, and no obligation of debt created. Any donations that occur from one person to the other are hypothesized to result from either unbounded altruism or warm-glow egoism (or the combination, impure altruism).[11] As such, the matching grant condition should have no effect on donations from separately-working participants.

H2: The matching grant will decrease sharing in a collaboratively-working condition.

In the collaborative condition, two people work on the same task in collaboration, with interdependent outcomes (each person's success depends on their partner also succeeding). If donations from one person to the other in this condition are motivated by a sense of debt owed to the collaborator, then a matching grant can decrease donations from the collaboratively-working participants. If the matching grant decreases donations in the collaborative condition, sharing in this condition cannot be explained by unbounded or

[11] These three mechanisms are not differentiable by way of a matching grant.

impure altruism or by warm-glow egoism.

H3: Collaboration will increase sharing in the absence of inequality.

If the act of collaborating itself creates a state of indebtedness to one's collaborator, the increased willingness to share earned resources that results from this debt of obligation should appear even when operating from baseline conditions of equality between the collaborating partners. In other words, inequality of outcomes should not be necessary to observe the prosocial effect of collaboration.

H4: Collaboration will not increase willingness to take from one's partner in order to move towards equality.

Moreover, if the effect of collaboration on willingness to sacrifice arises from a sense of something owed to the collaborator, then collaboration should not be expected to increase willingness to take from one's partner in an effort to equalize outcomes. Instead, by creating this bond of obligation, collaboration could be expected to *decrease* willingness to take.

V. Results

Experiments 1 & 2 - Bounded altruism vs. unbounded or warm-glow

Results of an initial test—Experiment 1—and replication—Experiment 2—are shown in Figure 2. Figure 2 plots mean amount given to the partner in collaboratively- and separately-working conditions, under standard giving (the recipient receives the amount donated by the participant) and under a matching grant (the recipient receives double the amount donated by the participant). Both experiments exhibit a collaboration effect under the standard giving condition: Collaboration increases the mean amount given to the partner compared to giving in the separately-working group, in Experiment 1 ($t(522) = +\$.07$, 95% CI: $\$.05, \$.10$, $d = .49$, $p = .000$) and Experiment 2 ($t(559) = +\$.05$,

95% CI: $\$.02, \$.07,\ d = .29,\ p = .001$).

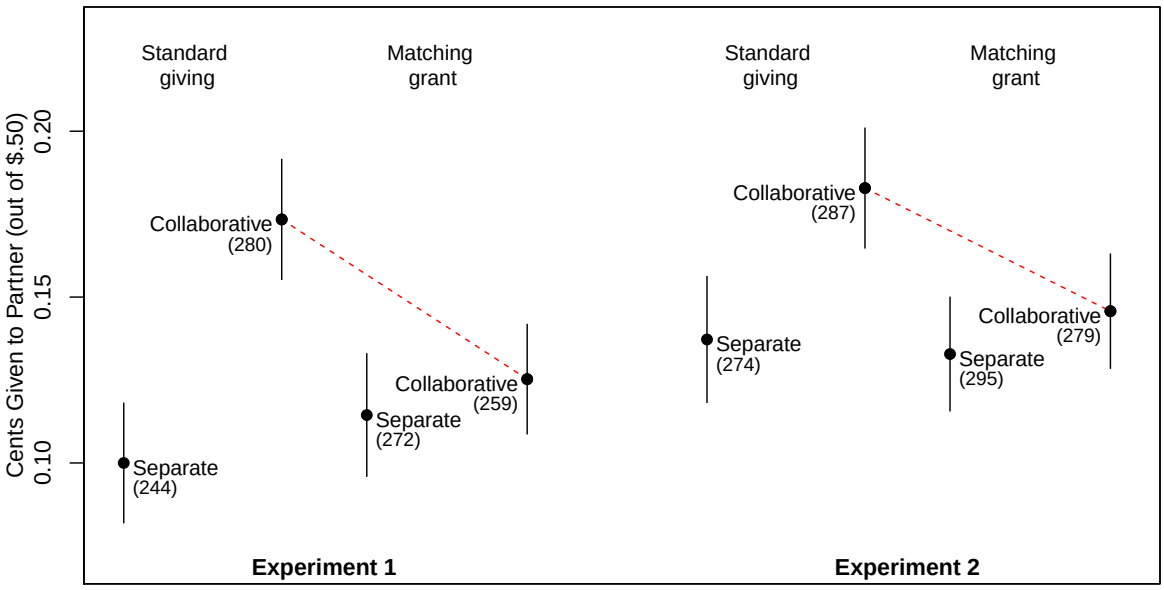

Figure 2

Mean amount given to the partner within each group, with 95% confidence intervals and n shown in parentheses. Results from Experiment 1 are shown on the left, from Experiment 2 on the right. Dotted lines highlight the decrease in giving caused by the matching grant in the collaboratively-working conditions. As predicted, there is no effect of the matching grant on giving in the separately-working conditions. Full distributions shown in SI Appendix 4.

In both experiments, the mean amount shared in the collaboratively-working group under standard giving is greater than $\phi/\kappa = \$.125$, the sufficient condition for a matching grant to decrease the amount shared (Exp 1: M=$.173, S.D. = $.156; Exp. 2: M=$.183, S.D. = $.158). In both experiments, as predicted the matching grant significantly decreases mean amount shared in the collaboratively working group (Exp 1: t(537)= $-\$.05$, 95% CI: $-\$.07, -\$.02,\ d = .33,\ p = .000$; Exp. 2: t(564) = $-\$.04$, 95% CI: $-\$.06, -\$.01,\ d = .24,\ p = .004$) and has no effect on mean amount shared in the separately working group (Exp 1: t(514)=+$.01, 95% CI: $-\$.01, +\$.04,\ d = .09,\ p = .212$;

Exp. 2: t(567) = −$.00, 95% CI: −$.03, +$.02, $d = .03$, $p = .738$). Table 1 shows that the estimated interaction between the collaboration treatment and the matching grant condition is statistically significant.

Table 1

Matching Grant Decreases Effect of Collaboration

Outcome: Amount given to partner	
Collaboration	$.06***
	[.01]
Matching Grant	$.00
	[.01]
Collab*Matching	−$.05***
	[.01]
Constant	$.11
	[.01]
N	2190

Fixed Effects for Experiment.
Robust SE shown in brackets.
*** indicates $p < .000$

Experiments 3 & 4 - Inequality aversion

Experiment 3 tests for an effect of collaboration in a context of baseline equality between the participant and the partner. Here, in both the collaboratively-working condition and in the separately-working condition, the participant and the partner receive the same amount of money for their work (in contrast to Experiments 1 & 2, in which the participant receives a windfall relative to the partner).

Results from Experiment 3 show that collaboration increases willingness to sacrifice when starting from a baseline of equality. When the participant and partner have received the same amount to begin with, collaboration nevertheless increases willingness to give some amount of that earned money to one's partner, moving from an average of $.06 (out

of \$.25 possible) given in the separately-working group to an average of \$.08 given in the collaboratively-working group (t(754) = +\$.02, 95% CI: +\$.00, +\$.04, $d = .15$, $p = .04$).

Note that the mean amounts given to the partner in Experiment 3 are considerably lower in both conditions (separately-working and collaboratively-working) than are the equivalent means in Experiments 1 & 2, which start from a baseline of inequality. However, the effect of collaboration in Experiment 3 represents a 30% increase over the average amount given in the separately working group. This effect of collaboration is of a similar magnitude to the effect shown under conditions of inequality (McGrath and Gerber, 2019).

Experiment 4 examines whether collaboration increases willingness to take from one's partner in order to achieve equality. In both the collaboratively-working and separately-working conditions in Experiment 4, participants receive a \$.20 bonus payment in the lottery while their partners receive a \$.40 bonus payment. In this experiment, participants are given the option to take any amount of the \$.40 bonus from their partner before the outcomes are finalized. Participants are informed that this decision is entirely anonymous: the partner will not know the participant was presented with this opportunity, and any amount remaining will be presented to the partner as if determined by the lottery.

If the effect of collaboration is simply to increase inequality aversion, it should increase willingness to take from one's partner in order to equalize the outcomes. Instead, collaboration exerts the opposite effect, decreasing willingness to take from the partner's surplus. In the separately-working group, participants take \$.10 on average, compared to \$.06 in the collaboratively-working group (t(266)=−\$.03, 95% CI: −\$.06, −\$.01, $d = .29$, $p = .02$).

VI. Discussion

Collaboration, even a minimal instance, is able to induce willingness to sacrifice for the sake of one's collaborator. Why does collaboration exert this effect? How does it operate? The evidence presented here indicates that collaboration generates a bounded

form of altruism. Experiments 1 & 2 show that collaboration does not appear to increase willingness to sacrifice from an internally-facing "warm-glow" egoism (controlling for other forms of egoism, such as reputation or future gain),[12] or out of a general desire to improve the other's welfare (i.e., an internalization of utility as can be expected with bonds of kinship). Instead, collaborators operate as if under an obligation of debt: what is due to a collaborator appears to be different from what is due to someone working just as hard, but working separately. The increased prosociality that results from collaboration does not reflect an increased benevolence, but rather an altered sense of equity.

Further insight into the effects of collaboration on perceptions of equity is provided by Corbit et al., 2017, who identify a novel form of inequity aversion that results from successful collaboration. The authors find evidence of increased willingness to destroy all resources (both partners receive nothing, 0:0) rather than accept an unfair advantage (4:1). Alongside the increased prosociality previously documented, Corbit et al.'s novel findings of an increased preference for detrimental equality provide important nuance to the ways in which collaboration changes our perception of what is fair.

But evidence that collaboration alters our sense of what is fair does not mean that the effect of collaboration amounts to or is driven by a desire to equalize outcomes. As Experiments 3 & 4 demonstrate, the prosocial effect of collaboration occurs even when outcomes are already equal, and collaboration *decreases* willingness to take from one's partner in order to equalize outcomes. Indeed, in both of these experiments, the increased willingness to sacrifice induced by collaboration manifests as a decrease in inequality aversion.

Collaboration has multi-faceted effects on our sense of what is fair. The prosocial effect of collaboration on willingness to sacrifice is not driven by a principlistic desire to equalize outcomes, nor is it a product of increased affinity or benevolence. Instead,

---

[12] Indeed, the results of Experiments 1 & 2 suggest a "calculating" nature to the sharing behavior that, if observed, could constitute a negative reputational signal (see Jordan et al., 2016)

collaboration appears to produce a complex, two-way bond of obligation.

Because the goal of the present investigation is to gain understanding of the effect of collaboration at a fundamental level, the findings here are based on a minimalist form of collaboration, pared to its essential elements. How the effects of collaboration vary under conditions reflecting different real-world scenarios is an important question for future investigation. The experiments I present in this study set the foundation for future work to systematically introduce additional variables of interest to evaluate how the results may change, adding nuance to the baseline established here.

These findings about how the collaboration effect operates at a fundamental level provide insight into the way in which collaboration alters our relationships with others. Strictly speaking, the prosocial effect of collaboration appears to constitute altruism: it generates a costly action that confers an advantage to another individual (Fehr & Fischbacher, 2003; Fowler & Kam, 2007), undertaken without regard to one's own future gain, and maintained under different circumstances (Experiments 1/2, 3, & 4) that vary the relationship between potential ultimate goals (Batson & Shaw, 1991).[13]

But the results reported here indicate that the willingness to bear costs that is induced by collaboration may represent altruism in its most limited form: an altruism of just what is due and nothing more. Rather than prompting a general benevolence or the internalization of another's well-being, in its most essential form the effect of collaboration appears driven by a somewhat miserly but deeply embedded sense of what is owed. At a fundamental level, we act generously toward our collaborators because we feel that we must.

Why does this distinction matter? While different motivations may generate the

---

[13] Batson & Shaw note, "we do not observe another person's goals or intentions directly; we infer them from the person's behavior. ...we can draw reasonable inferences about a person's ultimate goal if we can observe the person's behavior in different situations that involve a change in the relationship between the potential ultimate goals. The behavior should always be directed toward the true ultimate goal." (1991, p. 110).

same prosocial behavior in the immediate context, they can have very different implications down the line.[14] Appealing to the collaborator principle may induce prosociality, but—with its inherent limitedness—narrow people's focus to when they have fulfilled their obligation and owe nothing further.

[14] See, e.g., DellaVigna et al., 2012 for field-experimental evidence that voluntary sharing decisions made under social pressure can be welfare-reducing.

References

Andreoni, J. (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal*, *100*(401), 464–477. https://doi.org/10.2307/2234133

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390–1396. https://doi.org/10.1126/science.7466396

Bargh, J. A., & Morsella, E. (2008). The Unconscious Mind. *Perspectives on Psychological Science*, *3*(1), 73–79. https://doi.org/10.1111/j.1745-6916.2008.00064.x

Bartlett, M. Y., & DeSteno, D. (2006). Gratitude and prosocial behavior : Helping when it costs you. *Psychological Science*, *17*(4), 319–325. https://doi.org/10.1111/j.1467-9280.2006.01705.x

Bateson, P., & Laland, K. N. (2013). Tinbergen's four questions: An appreciation and an update. *Trends in Ecology and Evolution*, *28*(12), 712–718. https://doi.org/10.1016/j.tree.2013.09.013

Batson, C. D., Ahmad, N., Lishner, D. A., & Tsang, J.-A. (2002). Empathy and Altruism. In *Oxford handbook of positive psychology*. Oxford University Press.

Batson, C. D., & Shaw, L. L. (1991). Evidence for Altruism: Toward a Pluralism of Prosocial Motives. *Psychological Inquiry*, *2*(2), 107–122. https://doi.org/10.1207/s15327965pli0202_1

Bear, A., & Rand, D. G. (2016). Intuition, deliberation, and the evolution of cooperation. *Proceedings of the National Academy of Sciences*, *113*(4), 201517780. https://doi.org/10.1073/pnas.1517780113

Becker, G. S. (1974). A Theory of Social Interactions. *Journal of Political Economy*, *82*(6), 1063–1093.

Blake, P. R., McAuliffe, K., Corbit, J., Callaghan, T. C., Barry, O., Bowie, A., Kleutsch, L., Kramer, K. L., Ross, E., Vongsachang, H., Wrangham, R., &

Warneken, F. (2015). The ontogeny of fairness in seven societies. *Nature*, *528*(7581), 258–261. https://doi.org/10.1038/nature15703

Bolton, G. E., & Ockenfels, A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition. *The American Economic Review*, *90*(1), 166–193.

Bowles, S., & Gintis, H. (2004). The evolution of strong reciprocity: Cooperation in heterogeneous populations. *Theoretical Population Biology*, *65*(1), 17–28. https://doi.org/10.1016/j.tpb.2003.07.001

Bowles, S., & Gintis, H. (2013). Social preferences. In L. Bruni & S. Zamagni (Eds.), *Handbook on the economics of reciprocity and social enterprise*.

Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology*, *94*(1988), S95–S120.

Corbit, J., McAuliffe, K., Callaghan, T. C., Blake, P. R., & Warneken, F. (2017). Children's collaboration induces fairness rather than generosity. *Cognition*, *168*, 344–356. https://doi.org/10.1016/j.cognition.2017.07.006

Dahl, A., & Paulus, M. (2019). From Interest to Obligation: The Gradual Development of Human Altruism. *Child Development Perspectives*, *13*(1), 10–14. https://doi.org/10.1111/cdep.12298

DellaVigna, S., List, J. A., & Malmendier, U. (2012). Testing for Altruism and Social Pressure in Charitable Giving. *The Quarterly Journal of Economics*, *127*(1), 1–56. https://doi.org/10.1093/qje/qjr050

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*(6960), 785–791. https://doi.org/10.1038/nature02043

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, (August), 817–868.

FeldmanHall, O., Dalgleish, T., Evans, D., & Mobbs, D. (2015). Empathic concern drives costly altruism. *NeuroImage*, *105*, 347–356. https://doi.org/10.1016/j.neuroimage.2014.10.043

Fershtman, C., Gneezy, U., & List, J. A. (2012). Equity Aversion: Social Norms and the Desire to be Ahead. *American Economic Journal: Microeconomics*, *4*(4), 131–144. https://doi.org/10.1257/mic.4.4.131

Fong, C. M., Bowles, S., & Gintis, H. (2006). STRONG RECIPROCITY AND THE WELFARE STATE. In S.-C. Kolm & J. M. Ythier (Eds.), *Handbook of the economics of giving, altruism and reciprocity* (pp. 1439–1464). Elsevier B.V. https://doi.org/10.1016/S1574-0714(06)02023-9

Fowler, J. H., & Kam, C. D. (2007). Beyond the self: Social identity, altruism, and political participation. *Journal of Politics*, *69*(3), 813–827. https://doi.org/10.1111/j.1468-2508.2007.00577.x

Greenberg, M. S. (1980). A Theory of Indebtedness. In M. S. G. Kenneth J. Gergen & R. H. Willis (Eds.), *Social exchange: Advances in theory and research* (pp. 3–26). Plenum Press.

Hamann, K., Warneken, F., Greenberg, J. R., & Tomasello, M. (2011). Collaboration encourages equal sharing in children but not in chimpanzees. *Nature*, *476*(7360), 328–31. https://doi.org/10.1038/nature10278

Hawley, P. H. (2014). Evolution, Prosocial Behavior, and Altruism. In L. M. Padilla-Walker & G. Carlo (Eds.), *Prosocial development: A multidimensional approach*. Oxford University Press. https://doi.org/10.1093/acprof

Impett, E. A., Gable, S. L., & Peplau, L. A. (2005). Giving up and giving in: The costs and benefits of daily sacrifice in intimate relationships. *Journal of Personality and Social Psychology*, *89*(3), 327–344. https://doi.org/10.1037/0022-3514.89.3.327

Jordan, J. J., Hoffman, M., Nowak, M. A., & Rand, D. G. (2016). Uncalculating cooperation is used to signal trustworthiness. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(31), 8658–8663. https://doi.org/10.1073/pnas.1601280113

Karlan, D. S., & List, J. A. (2007). Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field. *The American Economic Review*, *97*(5), 1774–1793. https://doi.org/10.1257/aer.97.5.1774

Konow, J. (2010). Mixed feelings: Theories of and evidence on giving. *Journal of Public Economics*, *94*(3-4), 279–297. https://doi.org/10.1016/j.jpubeco.2009.11.008

Levine, D. K. (1998). Modeling Altruism and Spitefulness in Experiments. *Review of Economic Dynamics*, *1*(3), 593–622. https://doi.org/10.1006/redy.1998.0023

Malmendier, U., Te Velde, V. L., & Weber, R. A. (2014). Rethinking reciprocity. *Annual Review of Economics*, *6*, 849–874. https://doi.org/10.1146/annurev-economics-080213-041312

McGrath, M. C., & Gerber, A. S. (2019). Experimental evidence for a pure collaboration effect. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-019-0530-9

Moreland, R. L., & Zajonc, R. B. (1982). Exposure effects in person perception: Familiarity, similarity, and attraction. *Journal of Experimental Social Psychology*, *18*(5), 395–415. https://doi.org/10.1016/0022-1031(82)90062-2

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*(7063), 1291–8. https://doi.org/10.1038/nature04131

Peng, C., Nelissen, R. M. A., & Zeelenberg, M. (2018). Reconsidering the roles of gratitude and indebtedness in social exchange. *Cognition and Emotion*, *32*(4), 760–772. https://doi.org/10.1080/02699931.2017.1353484

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *The American Economic Review*, *83*(5), 1281–1302.

Reeve, J. (2018). *Understanding motivation and emotion* (7th ed.). John Wiley & Sons.

Salovey, P., Mayer, J. D., & Rosenhan, D. L. (1991). Mood and helping: Mood as a motivator of helping and helping as a regulator of mood. In M. Clark (Ed.), *Prosocial behavior*. Sage Publications, Inc.

SI. (2021). https://arch.library.northwestern.edu/collections/bz60cw536

Sobel, J. (2005). Interdependent Preferences and Reciprocity. *Journal of Economic Literature*, *43*(June), 392–436. http://qjmam.oxfordjournals.org/content/41/4/503.abstract

Sugden, R. (1984). Reciprocity: The Supply of Public Goods Through Voluntary Contributions. *The Economic Journal*, *94*(376), 772–787.

Tinbergen, N. (1963). On aims and methods of Ethology. *Zeitschrift für Tierpsychologie*, *20*(4), 410–433. https://doi.org/10.1111/j.1439-0310.1963.tb01161.x

Tomasello, M. (2020). The Moral Psychology of Obligation. *Behavioral and Brain Sciences*, *43*(e56), 1–58. https://doi.org/10.1017/S0140525X19001742

Tomasello, M., & Vaish, A. (2013). Origins of Human Cooperation and Morality. *Annual review of psychology*. https://doi.org/10.1146/annurev-psych-113011-143812

Tsang, J.-A. (2006). The Effects of Helper Intention on Gratitude and Indebtedness. *Motivation and Emotion*, *30*(3), 199–205. https://doi.org/10.1007/s11031-006-9031-z

Van Lange, P. A., Drigotas, S. M., Rusbult, C. E., Arriaga, X. B., Witcher, B. S., & Cox, C. L. (1997). Willingness to sacrifice in close relationships. *Journal of Personality and Social Psychology*, *72*(6), 1373–1395. https://doi.org/10.1037/0022-3514.72.6.1373

Warneken, F., & Tomasello, M. (2013). The emergence of contingent reciprocity in young children. *Journal of Experimental Child Psychology*, *116*(2), 338–350. https://doi.org/10.1016/j.jecp.2013.06.002

Watkins, P. C., Scheer, J., Ovnicek, M., & Kolts, R. (2006). The debt of gratitude: Dissociating gratitude and indebtedness. *Cognition and Emotion*, *20*(2), 217–241. https://doi.org/10.1080/02699930500172291