Two-worker competition in gift-exchange: assessing intention-based reciprocity and inequity aversion

Bogliacino, Francesco and Rodríguez González, Nicolás

Universidad Nacional de Colombia, Centro de Investigaciones para el Desarrollo

March 2020

Two-worker competition in gift-exchange: assessing intention-based reciprocity and inequity aversion*

Francesco Bogliacino†
Nicolás Rodríguez González‡

March, 2020

Abstract

In this article, we study a three-person gift exchange, where two workers compete for a bonus. We derive the equilibrium properties of the models of sequential reciprocity and inequity aversion. We then prove a comparative statics theorem, when one worker becomes more productive. We show that compared with the predictions of outcome based model, those of the intention based model contrast sharply. This creates an ideal setting in which to perform a controlled experiment to test them. Our results largely support sequential reciprocity.

Keywords: Gift exchange, sequential reciprocity, inequity aversion

1 Introduction

A well-documented finding in the study of decision-making is that subjects behave as if they were incorporating the welfare of other people into their objective function.

The need to assume other regarding motivations has led to the development of a vast body of theoretical literature, mainly along two lines: outcome based models, where subjects compare payoffs across agents; and intention based models, where subjects evaluate the "context" under which certain strategies are chosen. Two concepts are particularly important in this body of literature: one is the aversion towards inequality and the other is reciprocity. Inequity aversion is defined as a situation in which agents are willing to bear a cost to reduce inequality, both favourable and unfavourable. Reciprocity refers to a situation in which the intentions of the players are explicitly taken into account when choosing which course of action to follow, with the aim of rewarding those who intend to be kind.

The goal of this article is to assess the predictive power of reciprocity versus inequity aversion models.

*We thank Alexander Gotthard, Cagatay Kayı, Oscar Benavides for comments on Nicolás' Master thesis. We also thank Camilo Gómez, Rafael Charris, and Felipe Montealegre for suggestions and discussion. We appreciate comments from the participants to our session at the 2019 North American ESA Meeting, Los Angeles, to SABE session at WEAI in San Francisco, and to different seminars in Bogota. We thank Universidad Nacional for the funding (Internal grant, Convocatoria JAB 2017).

†Corresponding author: Universidad Nacional de Colombia and Centro de Investigaciones para el Desarrollo, Bogota. fbogliacino@unal.edu.co

‡Universidad Nacional de Colombia, Bogota. nirodriguezgo@unal.edu.co

We proceed in the following way. We study a situation that includes three key features: the number of players is greater than two, competition and cooperation coexist and perfect equality is not a feasible outcome. In particular, we study a gift exchange with two workers competing for a bonus. In this setting, we introduce preferences according to some standard models of inequity aversion and reciprocity and study the basic equilibrium properties. We then introduce a shock and prove a comparative statics theorem for the two models. The shock is a productivity increase that benefits only one worker. As we derived, the predictions are strikingly different between reciprocity and inequity aversion, and a controlled experiment is an ideal setup in which to assess the two theories. For the experiment, we use the following design. We run a three-person gift exchange in three steps; that is, choice of wage, choice of effort, and choice of bonus. Then, we introduce the shock in the form of making a randomly chosen worker more productive. Additionally, we introduce two further treatments, where the more productive worker is chosen by merit or need, which makes it more salient and can be interpreted as an intention to treat the manager's guilt or her sensitivity to reciprocity. Our results support reciprocity against inequity aversion.

Our paper is related to the theoretical literature on other regarding preferences. The two reference models we use are Fehr and Schmidt (1999) and Dufwenberg and Kirchsteiger (2004), which extends the original work of Rabin (1993). The former is the typical model where agents have distributional concerns, while the latter is explicitly modelling intentions.

Since there are few applications of reciprocity models, mainly due to its complicated structure and the difficulties of studying equilibria, the exploration of the reciprocity model in this setting is the first contribution of this paper.

In the literature, the assessment of models of other regarding preferences has been performed in various ways. One possibility is to calibrate the model and estimate the goodness of fit with respect to experimental data (Fehr & Schmidt, 1999). An alternative is within subject assessment, where the same subjects are asked to perform similar games, and the point estimates of the parameters from one task are used to make predictions in other tasks (e.g., Blanco, Engelmann, and Normann (2011) for the inequity aversion model). Yet another alternative is to try to nest the models and test the robustness of the restrictions. An example of this strategy is Charness and Haruvy (2002), for the altruism based versus reciprocity based versus inequity aversion based models.

Empirically, this paper relates to two broad sets of literature. The first is on gift exchange. The seminal work by Akerlof (1982) aimed to explain involuntary unemployment by adding a sociological justification for the unwillingness by a firm to cut wages, when the latter stood above the market clearing level. Since that milestone, much theoretical and empirical work has been conducted, which is beyond the scope of this article to summarize. Experimental evidence shows that the gift exchange is quite robust but its magnitude is very sensitive to the details of the game, and in experiments, to the settings (Charness & Kuhn, 2011). This paradigm is powerful enough to be able to understand the labor market, since it is simple and flexible at the same time. It is our claim that including more than two players, cooperation and competition, and the unfeasibility of perfect equality is important to be able to understand real world labor markets.

A second strand of related literature is that on the empirical assessment of other regarding preferences, which have increased our understanding of the functioning of modern economies. The existence of trust and reciprocity increases the set of contracts that are enforceable, and changes the prediction as to what type of institutions are best suited to solve social dilemmas.

Nevertheless, different models entail different predictions and the understanding of which representation of preferences fits better with specific choice problems is still an open question and requires further research.

The comparison between models of other regarding preferences raises three key questions: first, (external validity) how robust are the predictions to changes in settings; second, (conclusiveness) are there environments in which we can test these classes of models in a mutually exclusive way; and third, (degrees of freedom) to the extent that all of these models depend on a number of underlying unobservable parameters, when we collect data from treatments where a charged framing is introduced to affect those parameters, will these data be consistent with the predictions of the models?

This article's contribution to the three research questions is as follows. We study both outcome based models and intention based models in a setting in which we have more than two agents, sequential structure, competition and cooperation, and where perfect equality is not a feasible outcome, thus, contributing to the discussion on external validity. Although the prediction of sequential reciprocity is not easily extendable, we show that the conclusions can be generalized to a setting with weakened assumptions. Moreover, we test the two models in an alternative way, beyond aggregate or within subject assessment. Finally, in the experiment, we manipulate the rules of assignment as to who is the more productive as an intention to treat psychological parameters, and to the best of our knowledge, this is the first paper to do so.

This article proceeds as follows. Section 2 describes the theoretical framework. Section 3 presents the main predictions and Section 4 introduces the experimental design and the results. Section 5 concludes and discusses. All proofs are in the Appendices.

2 Theoretical framework

2.1 Game

Consider an interaction between two workers $i, -i$ (he/his) and one manager M (she/her), in which workers simultaneously choose an effort level out of the set $E = \{e_1, e_2, e_3\}$, where e_1, e_2 and e_3 represents low, medium and high effort, respectively. We assume that $0 < e_1 < e_3$ and $e_2 = \frac{e_1+e_3}{2}$. In the next stage, the manager observes the efforts and decides to give a fixed bonus b to one worker or none at all.

The payoff of the worker i related to the terminal nodes of the game is given by:

$$\pi_i(e^i, e^{-i}, (b_i, b_{-i})) = w + b_i - c(e^i),$$

where w is an exogenous wage, b_i is the salary bonus assigned to the worker (it can be 0 or b), e^i is the implemented effort by the worker, and c is an increasing and strictly convex function defined in a neighborhood that contains the interval $[e_1, e_3]$.

The payoff of the manager M is given by:

$$\pi_M(e^i, e^{-i}, (b_i, b_{-i})) = \sum_{k \in W} (e^k - b_k),$$

where W is the set of workers.

We make the following assumptions:

Assumption 1 (Profitability hypothesis). *For each worker i, we assume:*

1.a) $\pi_i(e_3, e^{-i}, (b, 0)) > \pi_i(e_1, e^{-i}, (0, b_{-i}))$, *and*

1.b) $\pi_M(e_3, e^{-i}, (b, 0)) > \pi_M(e_2, e^{-i}, (0, 0))$.

In other words, reciprocity is profitable.

Assumption 2 (Inequality Hypothesis). *For each worker i, we assume:*

$$\pi_i(e^i, e^{-i}, (b_i, b_{-i})) < \pi_M(e^i, e^{-i}, (b_i, b_{-i})),$$

for all $(e^i, e^{-i}, (b_i, b_{-i}))$. In every terminal node, the manager obtains a larger payoff than the workers.

The first assumption is necessary in order to make reciprocity meaningful. The second assumption is motivated by empirical relevance.

2.2 Basic models for reciprocity and inequity aversion

Although there are various models that allow for studying preferences with reciprocity and inequity aversion, we have focused on Dufwenberg and Kirchsteiger (2004) and Fehr and Schmidt (1999) (hereafter DK and FS). We chose the former because it is a generalization of Rabin's approach (1993) for a dynamic context. Moreover, this model only has the sensitivity to reciprocity as a free parameter, while Falk and Fischbacher (2006) have both reciprocity parameters and intention factors.

Finally, although we acknowledge the contribution by Çelen, Schotter, and Blanco (2017), their model has some additional degrees of freedom that makes it harder to derive refutable predictions for specific settings. Levine (1998) is also relevant, but models types (altruist and spiteful) instead of intentions. Segal and Sobel (2007) can be considered a foundational paper that provides axiomatic bases for the representation of reciprocity, including the one used in this paper.

On the other hand, we chose FS over Bolton and Ockenfels (2000), mainly because of its more widespread use.

2.2.1 Inequity aversion model

We adopt the FS model in its basic form. In this model, the agents get disutility in two forms from inequality: they feel guilty for having more than others, and they feel envious for having less than others. Moreover, the intensity of the guilt and envy is heterogeneous among agents. Formally, in an interaction between n agents, where the profile of material payoffs is $\tilde{\pi} = (\tilde{\pi}_i)_{j=1}^n$ and $\tilde{\pi}_i$ represents the material payoff for the agent i, the inequality averse utility function[1] of agent i is:

$$\tilde{u}_i(\tilde{\pi}) \equiv \tilde{\pi}_i - \frac{\alpha_i}{n-1} \sum_{j \neq i} \max\{\tilde{\pi}_j - \tilde{\pi}_i, 0\} - \frac{\beta_i}{n-1} \sum_{j \neq i} \max\{\tilde{\pi}_i - \tilde{\pi}_j, 0\},$$

where $0 \leq \beta_i < 1$ y $\beta_i \leq \alpha_i$. The first inequality is a theoretical restriction justified because it seems reasonable that agents are not willing to burn one monetary unit in order to reduce the

[1]The notation $\tilde{u}$ represents the utility obtained by an averse inequity individual, in the FS sense.

advantageous inequality. From the construction of the utility function, the parameters α_i and β_i are known as envy and guilt, respectively; thus, the second inequality means that the agents dislike the disadvantageous inequality more than the advantageous one.

2.2.2 Reciprocity model

In our reciprocity framework, we construct a *psychological game*[2] in the current setting, in line with the DK model. First of all, the reciprocity theory posits that an agent responds to the level of kindness in others' actions (be nice to those who behave nicely and be unkind to those who have been unkind). Hence, an agent owns a belief about the others' strategies, in order to classify her own action as kind or unkind.[3] On the other hand, since reciprocity consists of a behavioral response to the kindness of others, the agent also needs a belief about the latter. In short, we need to include beliefs about the others' strategies (first order beliefs) and beliefs about the beliefs about those strategies (second order beliefs). Thus, provided that A_i represents i's strategy set, B_{ij} the beliefs of i about j's strategy set and C_{ijk} the beliefs of i over the beliefs of j about the k's strategy set, then it must be that $B_{ij} = A_j$ and $C_{ijk} = B_{jk} = A_k$. Let H be the finite set of histories of the game, then given $a_i \in A_i$ and $h \in H$, we define $a_i(h)$ as the updated strategy that prescribes the same choices as a_i, except the ones for the history h that are interpreted to occur with probability one. Similarly, we define $b_{ij}(h)$ and $c_{ijk}(h)$ for $b_{ij} \in B_{ij}$ and $c_{ijk} \in C_{ijk}$. The above allows agents to interpret the observed decisions as intentioned ones. In addition, we assume that in case agent j makes a decision after agent i, the belief of i about the belief of j over the action of i corresponds to the actual action taken by i; in other words, each player knows that his action will determine the belief of the players who make choices after him. This is mainly to rule out "weird" equilibria. For instance, a first mover can believe that the other believes that he will be unkind, so he thinks the other will take an unkind action in the next stage; and thus, he chooses an unkind action.[4]

To measure the kindness towards another agent, we add a norm of fairness[5] defined by:

$$EV_i[\hat{\pi}_j]((b_{ij})_{j\neq i}) \equiv \frac{1}{2}\left[\max_{a_i\in A_i} \hat{\pi}_j(a_i, (b_{ij})_{j\neq i}) + \min_{a_i\in A_i} \hat{\pi}_j(a_i, (b_{ij})_{j\neq i})\right]. \tag{1}$$

Observe that this norm depends on the beliefs of the agent about the counterparts' actions and corresponds to an expected payoff. Thus, the kindness of i with j is the "deviation" of the j's payoff with respect to the norm of fairness. Formally:

$$\kappa_{ij}(a_i(h), (b_{ij}(h))_{j\neq i}) \equiv \hat{\pi}_j(a_i(h), (b_{ij}(h))_{j\neq i}) - EV_i[\hat{\pi}_j]((b_{ij}(h))_{j\neq i}). \tag{2}$$

[2] For a definition, see Geanakoplos, Pearce, and Stacchetti (1989). They offer an extended framework (psychological game theory) to overcome the limitation of standard game theory in order to model the psychological dimension of strategic interactions. Nevertheless, their framework has some limitations for extensive-form games. Battigalli and Dufwenberg (2009) provide an extension.

[3] For instance, suppose you are thinking of giving a chocolate to a friend, but you do not know whether or not your friend has started a diet. If he has, giving the chocolate is an unkind action, but if he has not, giving it is a kind action.

[4] These 'self-fulfilling expectations' equilibria are a characteristic result of DK. The introduction of the mentioned condition is sufficient for eliminating such equilibria. It is also very plausible.

[5] Within the interaction analyzed, this norm of fairness is equivalent to the equitable payoff defined by DK, which has generated some discussion. In particular, Isoni and Sugden (2019) show that the use of this concept does not explain the emergence of trust. Nonetheless, Dufwenberg and Kirchsteiger (2019) show that this could be the case by slightly changing the definition of the equitable payoff. That variation is immaterial in our setting.

Now, according to the reciprocity theory, the agent i chooses a kind action towards j if and only if j has chosen a kind action towards i. For that reason, it is necessary to define the belief of i about the degree of kindness by j towards i. Such belief is defined by:

$$\lambda_{iji}(b_{ij}(h),(c_{ijk}(h))_{k\neq j}) \equiv \hat{\pi}_i(b_{ij}(h),(c_{ijk}(h))_{k\neq j}) - EV_j[\hat{\pi}_i]((c_{ijk}(h))_{k\neq j})). \quad (3)$$

With the previous definitions, it is possible to construct a utility function contingent upon the histories of the game that represents preferences for reciprocity. Such function is defined by:[6]

$$\begin{aligned}\hat{u}_i(a_i(h),(b_{ij}(h),(c_{ijk}(h))_{k\neq j})_{j=i}) &\equiv \hat{\pi}_i(a_i(h),(b_{ij}(h)) \\ &+ \sum_{j\in N-\{i\}}(\gamma_{ij}\kappa_{ij}(a_i(h),(b_{ij}(h))_{j\neq i})\lambda_{iji}(b_{ij}(h),(c_{ijk}(h))_{k\neq j})),\end{aligned} \quad (4)$$

where γ_{ij} measures how important it is for i to establish reciprocal relations with j. We name this the sensitivity to reciprocity by i towards j. Notice that the *psychological utility* of each player is the sum of his material payoffs and the psychological payoffs associated with every player.

Thus, we now move to define the equilibrium concept.

Definition 1. *A profile of strategies $a^* = (a_i^*)_{i=1}^n$ is a sequential reciprocity equilibrium (SRE) if for all $i \in N$ and each $h \in H$ it satisfies:*

1. $a_i^*(h) \in argmax_{a_i\in A_i(h,a^*)}\hat{u}_i(a_i,(b_{ij}(h),(c_{ijk}(h))_{k\neq j})_{j=i})$,
2. $b_{ij} = a_j^*$ *for all* $j \neq i$, *and*
3. $c_{ijk} = a_k^*$ *for all* $j \neq i$, $k \neq j$.

Notice that the equilibrium concept relies on the assumption that the beliefs are correct in equilibrium and that every player knows the sensitivity to reciprocity of each player. The latter assumptions may appear unrealistic,[7] and as a consequence, we explore the implications of introducing uncertainty, showing that the main results are unaffected (see *infra*).

3 Main predictions

3.1 Reciprocity

First, let us introduce some mathematical notation. Take $i \in W$ to be a worker and j, k any type of agent, then:

- $p_{1i} \equiv \text{Prob}(e^i = e_1)$, $p_{2i} \equiv \text{Prob}(e^i = e_2)$ and $p_{3i} \equiv \text{Prob}(e^i = e_3)$. Also, $\sigma_i \equiv (p_{1i}, p_{2i}, p_{3i})$,
- $e \equiv (e^j)_{j\in W}$. Moreover, $r_i(e) \equiv \text{Prob}(b_i = b|e)$ and $r(e) \equiv ((r_j(e))_{j\in W}, 1 - \sum_{j\in W} r_j(e))$. Observe that $r_i(e)$ is the probability that the worker i gets the bonus in the history determined by e. Thus, one manager's behavioral strategy can be written by $r = (r(e))_{e\in E\times E}$, in other words, a behavioral strategy for M is a combination of probability distributions over her actions in each possible game history,

[6]The notation with $\hat{u}$ is used to represent the psychological utility function for an agent motivated by reciprocity.

[7]We are grateful to Çağatay Kayı, who pointed out this limitation of the model to us.

- p_{ni}^{j} is the belief of j over p_{ni}. Likewise is defined $r_i^j(e)$, σ_i^j, $r^j(e)$ and r^j, and
- p_{ni}^{jk} is the belief of k over p_{ni}^{j}. Likewise is defined $r_i^{jk}(e)$, σ_i^{jk}, $r^{jk}(e)$ and r^{jk}.

Since the manager is the second mover, she takes the workers' choices as intentioned. Therefore, she responds directly to the degree of kindness of the workers' actions. Consequently, the following claim holds:

Claim 1. *Let be $i \in W$ and $e \in E \times E$. If $e^i \neq e_3$, then the manager chooses $r_i(e) = 0$. In other words, when a worker does not choose the highest effort, he will not get the bonus.*

Proof. See Appendix A □

In equilibrium, a worker is not granted the bonus unless he chooses the high effort. This result is due to the fact that low and medium effort are not kind actions. Moreover, even when a worker implements the high effort, the manager could decide not to give the bonus to him; in other words the following claim holds:

Claim 2. *Let be $i \in W$ and $e \in E \times E$. In equilibrium, if $e^i = e_3$, and:*

1. *$e^{-i} \neq e_3$ and $\gamma_{Mi} > \frac{1}{e_3 - e_2}$, the manager chooses $r_i(e) = 1$. In other words, if a worker chooses the highest effort, his co-worker does not, and the manager is sensitive enough to reciprocity, the worker with the best performance gets the bonus,*
2. *$\gamma_{Mi} < \frac{1}{e_3 - e_2}$, then the manager chooses $r_i(e) = 0$. In other words, even if the worker has chosen the highest effort, the manager does not give the bonus to him when she is not sensitive enough to reciprocity, and*
3. *$e^{-i} = e_3$ and $\gamma_{M-i}, \gamma_{Mi} > \frac{1}{e_3 - e_2}$, the manager gives the bonus to the worker for whom she feels a greater sensitivity to reciprocity.*

Proof. See Appendix A □

Thus, the manager gives the bonus to a worker on the condition that she is sensitive enough to reciprocity and the worker has chosen the high effort. When two workers choose the high effort, the manager gives the bonus to whom she feels more reciprocal.

On the other hand, the workers can anticipate the manager's optimal behavior, so if one of them thinks that he will not receive the bonus, even when he chooses to strive, he will implement the low effort. Formally, we can prove the following claim:

Claim 3. *Let be $i \in W$. In equilibrium, if $r_i^i = 0$, then i chooses $p_{1i} = 1$. In other words, if a worker believes that the manager will not give the bonus to him regardless of his choice, he implements the lowest effort. In particular, whether $\gamma_{Mi} < \frac{1}{e_3 - e_2}$ (Claim 2.2.) or $\gamma_{Mi} < \gamma_{M-i}$ and $p_{3-i}^{i} = 1$ (Claim 2.3.).*

Proof. See Appendix A □

The previous result characterizes the worker's behavior in situations of negative reciprocity, namely when the manager is not sensitive enough to reciprocity with him, or when the manager is sensitive enough towards him, but she is more sensitive to reciprocity towards the other worker, and the latter chooses the high effort.

In addition, each worker knows that his action determines the manager's belief because of the sequential structure. Thus, his expectations about the kindness of the manager are based upon his own action, which entails that if the manager is sensitive enough and this sensitivity is greater for the worker than towards his co-worker, he will get the bonus as long as he makes the maximum effort. As a result, under the Assumption 1, the following claim holds:

Claim 4. *Let be $i \in W$. If $r_i^i = 1$, then $p_{3i} = 1$. This means that if a worker believes that he will get the bonus when he chooses the highest effort, then he will do so. In particular, whether $\gamma_{Mi} > \frac{1}{e_3 - e_2}$ and $\gamma_{Mi} > \gamma_{M-i}$ (Claim 2.3.) or $\gamma_{Mi} > \frac{1}{e_3 - e_2}$ and $p_{-i3}^i = 0$ (Claim 2.1.).*

Proof. See Appendix A □

Observe that this result characterizes the worker's behavior for the following cases: when the manager is sensitive enough and her sensitivity is greater towards the worker than his co-worker; or when her sensitivity is greater towards the other but the latter does not choose the high effort.

Notice that the previous results fully characterize the optimal behavior under complete information, since the agents need to know the others' sensitivity to reciprocity. Therefore, we now move to the incomplete information case.

In order to do so, we start from the fact that the manager observes the workers' choices when she selects her action, responding to the kindness of the workers' actions, regardless of the beliefs she may have about their sensitivity to reciprocity. As a consequence, the manager's optimal behavior does not change under private information.

Conversely, the workers should anticipate the manager's behavior, thus their actions depend on the beliefs about her sensitivity to reciprocity. Since γ_{Mi}, $\gamma_{M-i} \geq 0$, we will assume that there exists a joint probability density function $f(\gamma_{Mi}, \gamma_{M-i})$ defined over $\mathbb{R}_+^2$. We will denote by $F(\gamma_{Mi}, \gamma_{M-i})$ the associated joint cumulative distribution function. It is important to point out that it is not reasonable to assume that γ_{Mi} and γ_{M-i} are independently distributed, because these parameters are associated with the same manager.

Finally, the worker's problem is to maximize his expected utility, and the following claim holds:

Claim 5. *Worker i chooses the high effort if and only if the probability of getting the bonus is large enough, specifically, if this probability is higher than $\frac{c(e_3) - c(e_1) + \frac{b\gamma_{iM}}{2}(e_3 - e_1)}{b + \frac{b\gamma_{iM}}{2}(e_3 - e_1)}$. Otherwise, he chooses the lowest effort.*

Proof. See Appendix A □

3.2 Inequity aversion

Remember that from this theoretical perspective, the agents get disutility in two forms from inequality: they feel guilty for having more than others, and they feel envious for having less than others.

Therefore, the optimal choice for a worker is the low effort, because otherwise he transfers resources to the richest agent in the interaction (Assumption 2), increasing the distributive gaps. Formally:

Claim 6. *In equilibrium, each worker implements the lowest effort.*

Proof. See Appendix B □

On the other hand, an inequity averse manager will reduce the inequality by giving the bonus to any worker in the case that she feels guilty enough.

Claim 7. *The manager gives the bonus to any worker if and only if her inequity aversion is large enough. Specifically, if $\beta_M > \frac{2}{3}$.*

Proof. See Appendix B □

An interesting implication of these results is that the optimal behavior of each agent is independent of the others' choices, since each agent has a strictly dominant strategy. Consequently, the availability of information about the inequity aversion parameters is irrelevant in equilibrium. In short, the theoretical predictions can be generalized from complete to incomplete information.

3.3 Main theorem

Observe that the former predictions depend on unobservable psychological parameters, such as guilt, envy and sensitivity to reciprocity. Therefore, we can always explain the behavior ex post, calibrating the parameters. As a result, we prefer to follow an alternative route: the construction of a comparative statics theorem which provides testable predictions about the outcome variable when we modify one observable parameter, *ceteris paribus*.

In order to construct the theorem, we start from the fact that the manager's decision depends on the workers' productivity under reciprocity theory. Specifically, we know that the manager gives the bonus on the condition that she is sensitive enough to reciprocity with the worker, which in this context means that $\gamma_{Mi} > \frac{1}{e_3 - e_2}$. Therefore, the threshold in terms of sensitivity to reciprocity to give the bonus is decreasing in the high effort. The above holds because the manager's payoff increases with e_3, so the maximum effort choice induces a stronger feeling of kindness (represented in the psychological payoffs). Thus, the manager no longer needs the same threshold of sensitivity to reciprocity to give the bonus, when e_3 increases.

Conversely, under inequity aversion, the productivity shock does not change the agents' behavior, since the productive worker always chooses the low effort to avoid increasing the distributive gaps between the manager and himself, and the latter will be willing to give the bonus to any worker in order to reduce inequity when her guilt is large enough.

The following theorem holds:

Theorem 1. *Let be $j \in W$ and e_{nj} the value of the effort e_n for worker j. If e_{3i} increases (i.e., worker i suffers a productivity shock) and the agents are motivated by reciprocity, then the probability of getting the bonus is increasing, conditional to worker i's high effort choice. Also, the probability that worker i implements the high effort increases by the productivity shock. In the case where the agents are motivated by inequity aversion, these probabilities are invariant.*

Proof. See Appendix C □

3.4 Extensions

What happens when we weaken some of the assumptions? The results turn out to be quite general. Indeed, the predictions of inequity aversion hold for an arbitrary number of workers, inasmuch as their behavior is independent of others' behavior (see Appendix B, Claims 6 and 7).

What about reciprocity? Intention-based models have been the subject of critiques for their limited scope for generalization. However, our sequential reciprocity claims also hold for an arbitrary number of workers. This is due to the fact that it is impossible to establish reciprocal interactions among workers (see Appendix A, Claim 3). Therefore, the only relevant interaction is between the manager and each worker, which means that the number of workers does not affect the predictions.

Another standard critique of the reciprocity model is the dependence on a universal and exogenous norm of fairness in order to classify actions as kind or unkind. Nonetheless, comparative statics holds for any norm of fairness defined as a strict convex combination between the maximum and minimal payoff.

Further, the theorem holds for heterogeneous norms of fairness.[8] This happens because high effort is always a kind action. Therefore, if the worker faces a productivity shock, high effort becomes kinder. Thus, *ceteris paribus*, it increases the likelihood of being rewarded by the manager. As an example, it may happen that workers expect more kindness from their manager than from their coworker. Even for this case, the theorem holds.

4 Experimental design and results

4.1 Design and procedures

The stage game of the experiment is as follows. There are three players, labelled worker I, worker II, and the company. In the instructions, we use the Experimental Currency Unit (ECU) as currency and we introduce a labor market framing.

The interaction occurs in three steps. In step one, the company chooses a discrete wage w in the set $[10, 34]$. Step one is added to allow for comparability with existing gift exchange experiments. In step two, the workers simultaneously choose the level of effort among five levels, with associated costs $(0, 1, 3, 6, 10)$. Five levels instead of three are added to determine robustness. Finally, in step three, the company decides whether to pay the bonus to either one or none of the workers.

The payoffs are computed according to the following formulas:

$$\begin{aligned} \pi_M &= 120 - 2w + \sum_{t \in W} (e^t - b_t) \\ \pi_i &= w + b_i - c(e^i). \end{aligned} \tag{5}$$

The initial endowment for the company is 120, which guarantees that Assumption 2 holds. Other parameters satisfy theoretical restrictions.

[8] The proof of this generalization is in Appendix D.

Participants go through five rounds of the same task. Matching is random and is never repeated on two consecutive rounds. Role assignment occurs at the beginning of the task and is maintained during the session.

We have four treatments. In the control condition, the interaction is repeated five times. In the three treatments, we introduce a productivity shock (i.e., worker two is assigned higher productivity for a given cost of effort). The difference among the three is the assignment criteria: in treatment one (hereafter, random treatment), worker 2 is randomly chosen; in treatment two (hereafter, socioeconomic strata, SES treatment), the participant who lives in the poorest neighborhood is chosen; and in treatment 3 (hereafter, productivity treatment) the most productive in a Real Effort Task is chosen (we use the task of summing two digits numbers, by Niederle and Vesterlund, 2007). Assignment to treatment is between-subjects at the session level.

In all sessions, the productivity associated with each level of effort for worker I is $4, 10, 17, 23, 30$; while the productivity associated with each level of effort for worker II is equal to workers I in control sessions, and is $4, 12, 20, 28, 36$ in the treatment sessions.

Sessions were run at the Universidad Nacional de Colombia in Bogotá. Participants were undergraduate students from different majors. All of the sessions had the same lead experimenter; one of the authors who is a native Spanish speaker. We ran a total of 12 sessions: four controls, three under treatment one, two under treatment two, and three under treatment three.

The experiment was programmed in the software zTree (Fischbacher, 2007). Each session had 15 participants, except for one control session with 12 participants. One ECU corresponds to 500 COP (0.15 USD). Each session lasted approximately 90 minutes, including the post-experimental questionnaire and the payment procedures. Average payment was 27 thousand COP (8.2 USD). The average hourly minimum wage in Colombia at the time was 1.05 USD. Experimental sessions were performed between October 2018 and February 2019.

4.2 Analysis

Our analysis plan includes the following steps: we first test for the presence of gift exchange, to check external validity with respect to other laboratory experiments. Second, we assess to what extent data are consistent with theoretical claims, with particular emphasis on reciprocity theory, which is our main contribution. Third, we test the main predictions of the comparative statics theorem, by pooling treatments together. Finally, interpreting the manipulation over the criteria of assignment as an intention to treat the underlying psychological parameters, we provide further tests of both theories.

Regarding the main theorem, under sequential reciprocity, the probability for worker II to choose the highest level of effort should be higher in treatments than control, and the probability to reward worker II, conditional on high effort, should be higher under treatments than under control. Conversely, for FS, these choices should be invariant to the productivity of the worker for the firm.

To assess these predictions, we run two regressions. The outcome variables are: the likelihood to choose the high effort by worker II (level four and five in our setting); and a dummy equal to one if worker II receives the bonus. We first assess the difference in mean between pooled treatments and control, by running OLS regressions with robust standard errors. If the effect of the treatment dummy is positive and significant, we can exclude inequity aversion as being the main driver behind observed behavior. Conversely, inequity aversion holds on the condition that the effect is not statistically different from zero.

We then estimate diff-in-diff regressions using the interaction between the treatment and a dummy equal to one if the worker II chooses the highest effort. The coefficient of the interaction should be positive for the reciprocity theory to be consistent with data, and zero for the theory of inequity aversion.

Next, we analyze data distinguishing among assignment criteria. In particular, SES and productivity treatments are intentions to shock the unobservable preference parameters. They can be interpreted as *fair* assignment criteria, making the company more willing to reciprocate a higher effort. Similarly, under a shock to guilt parameter (assuming that the manager becomes more inequity averse because we make more salient that one of the workers deserves to be more productive), the likelihood to reward one of the workers is non decreasing but not directed towards any particular worker, while the likelihood to choose a higher effort should be invariant to the fairness of the treatment.

To discriminate between the two predictions, we can estimate a diff-in-diff regression of the probability to assign a bonus to worker II in function of a dummy equal to one if the worker II chooses the highest effort, a dummy for fair treatment, and the interaction. Assume for a moment that the take up rate is positive; that is, the company is positively affected in its guilt parameter or reciprocity parameter. The coefficient for the dummy for fair treatment is the difference in the likelihood that worker II receives a bonus under fair treatment with respect to control, and should be non-negative if inequity aversion holds, while the interaction should not be statistically significant. Conversely, under sequential reciprocity, the dummy for fair treatment should be not significant and the interaction should be statistically significant. Of course, the assumption on the take up rate should be dealt with in the interpretation of the results.

4.3 Results

First, in Table 1, we report the main descriptive statistics and check the balancing of characteristics according to treatments. Results are consistent with random assignment.

Result 1. *The behavior is inconsistent with selfishness.*

If agents were selfish, the wage, effort and bonus would be the lowest, because they are costly to the agents. Nevertheless, the average wage is around 54 in a scale [0,100], where the lower and upper bounds represent the minimum and maximum wage, respectively. Also the modal effort is the highest one, and the bonus frequency is around 40%. These results are statistically significant at 5% level.

Table 1: Participant's descriptive statistics and Kruskal–Wallis Test for observable variables with respect to treatments

	Mean	Mode	Chi squared	P value
Sex	0.63	Male	1.69	0.64
Age	21.69	19 (15.25%)	1.92	0.58
Socioeconomic Strata	2.88	3 (47.46%)	1.97	0.57
Major		Economics (14.12%)		
RET score	3.82	4 (21.47%)	1.26	0.73
Term	5.42	8 (12.43%)	2.84	0.42
Locality		Suba (14.12%)		
Civil state		Single (95.48%)	0.031	0.99
Father's education		Bachelor (33.33%)	1.45	0.69
Mother's education		Bachelor (33.20%)	0.71	0.87

The null hypothesis is that all of the samples are from the same population. The χ^2 statistic has three degrees of freedom.

Result 2. *There exists a positive relation between wage and effort and between effort and bonus (extended gift-exchange).*

In Table 2, we observe the Spearman's rank correlation between effort and payment by the manager, which is positive and statistically significant. This suggests that the workers who received higher wages implemented a higher effort, and that the managers more frequently rewarded their hard work. Thus, we observe the gift-exchange phenomenon, even when two-workers compete for a bonus.

Table 2: Spearman's rank correlation coefficient for wage, effort and bonus by treatment

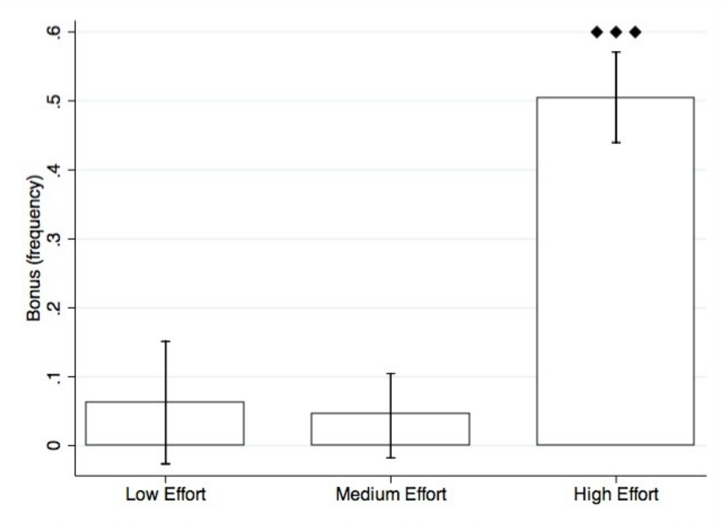
	Control		Random		Needs		Productivity	
	Worker I effort	Worker II effort	Worker I effort	Worker II effort	Worker I effort	Worker II effort	Worker I effort	Worker II effort
Wage	0.44***	0.34***	0.3***	0.09	0.13	0.49***	0.36***	0.5***
Worker I bonus	0.45***		0.35***		0.25*		0.49***	
Worker II bonus		0.41**		0.38***		0.4***		0.52***

*, **, *** represents significance at 10, 5 and 1%, respectively.

Result 3. *The behavior of the agents does not appear to be motivated by inequity aversion. The behavior of the manager is consistent with sequential reciprocity.*

Recalling that under inequity aversion the workers should choose the lowest effort, we can conclude that inequity aversion is rejected by Result 1 (the modal effort is the highest one). Furthermore, an inequity averse manager should take advantage of the bonus to decrease distributive gaps (that is, the bonus choice should be independent of the workers' performance). However, Figure 1 shows that instead, the bonus seems to be a mechanism to reward the workers' effort, which is consistent with reciprocity.

Figure 1: Bonus frequency conditional on the workers' effort



The reported effort corresponds to the maximum effort
of the workers' matched with one manager.
*** p<0.01,** p<0.05, * p<0.1.

Result 4. *The evidence on the effect of the productivity shock is consistent with sequential reciprocity.*

In Figure 2, we plot the frequency of the high effort choice of the worker II by control and treatments and we observe a tendency consistent with reciprocity in the probability of the high effort choice by the productive worker, although it is not statistically significant (Table 3). Nevertheless, we conduct a one sided binomial test that the frequency of the variable in the treatment is the same as control and the p-value is 0.051.

Figure 2: High effort frequency for worker II by shock

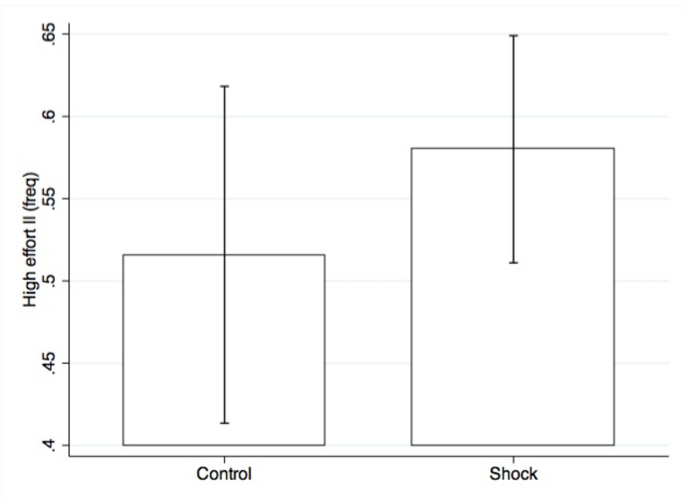


Table 3: OLS regression high effort choice of worker II in function of productivity shock

	Coefficient
Constant	0.52***
	(0.051)
Shock	0.064
	(0.062)
Obs	295

Robust standard errors in parenthesis.*** p<0.01, ** p<0.05, * p<0.1.

Additionally, as explained in the analysis plan, we estimate a diff-in-diff regression in which the outcome variable is the dummy equal to one if worker II receives the bonus, using the interaction between the treatment and a dummy equal to one if worker II chooses the highest effort. The results are shown in Table 4 and Figure 3. The coefficient for the choice of high effort should be interpreted as the increase in the likelihood of being rewarded when choosing high effort in the control condition, whereas the interaction coefficient measures the difference between the change in the likelihood of being rewarded by choosing the high effort under productivity shock, with respect to the change in the likelihood of being rewarded when making the same choice in the control condition.

Figure 3: Bonus towards worker II by effort and shock

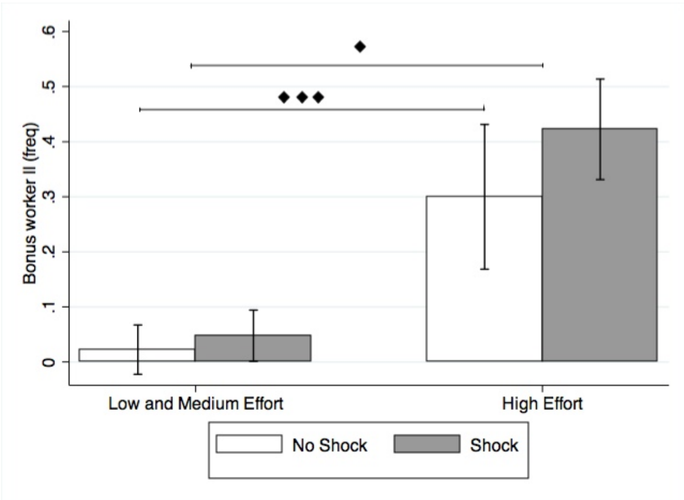


Table 4: Diff-in-diff regression worker II bonus, in function of high effort, productivity shock and interaction

	Coefficient
Constant	0.022 (0.022)
High effort	0.277*** (0.042)
Shock	0.025 (0.024)
High effort * Shock	0.097* (0.053)
Obs	295

Bootstrapped standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Therefore, given that the behavior of an inequity averse manager is not dependent on workers' actions, the above test rejects this prediction.

On the other hand, according to sequential reciprocity, we should have significance in the high effort coefficient, because the bonus is granted only under the high effort choice. Additionally, the interaction coefficient should be significant because the hard work is kinder on the part of the productive worker. This means that the good performance of a productive worker is more rewarded than the good performance of a standard worker, as reciprocity predicts in the main theorem. Thus, the experimental data support reciprocity.

Result 5. *The evidence of SES and productive treatment is consistent with sequential reciprocity.*

Finally, in Table 5 and Figure 4, we report and plot the results of a regression of the probability of paying a bonus to worker II over a dummy for high effort, a dummy for fair assignment (pooled SES and productivity treatments), and an interaction term. Here we use random treatment as the control, to avoid having a confounding factor.

Figure 4: Bonus towards worker II by effort and fairness

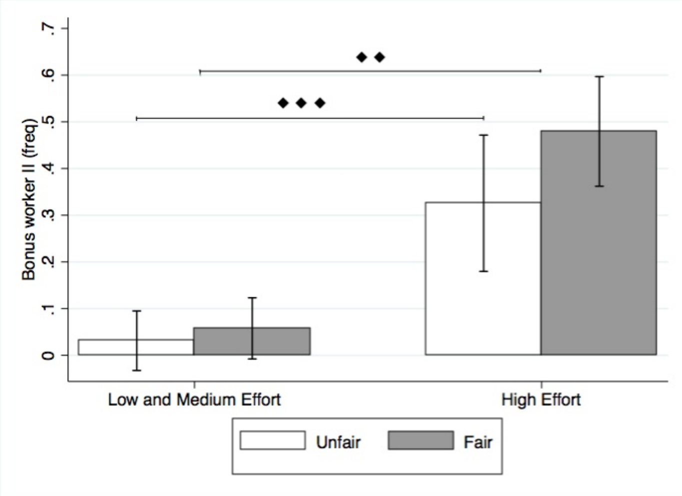


Table 5: Diff-in-diff regression worker II bonus, in function of high effort, fair treatments and interaction

	Coefficient
Constant	0.0312
	(0.028)
High effort	0.294***
	(0.028)
Fair	0.026
	(0.033)
High effort * Fair	0.127**
	(0.05)
Obs	200

Bootstrapped standard errors in parenthesis. *** $p<0.01$ ** $p<0.05$, * $p<0.1$.

If the agents were inequity averse, one could expect that the assignment criteria would make the manager unconditionally more culpable. Consequently, one would observe a positive (and significant) coefficient of the fair treatments dummy, because this represents the change in the likelihood of getting the bonus between random and fair assignment when the worker does not choose the high effort.

Conversely, under sequential reciprocity, one could expect that the assignment criteria would make the manager more sensitive to reciprocity towards the productive worker (he is more productive because he deserves it or he needs it). Therefore, we should have a positive and significant coefficient for the high effort dummy. The same should also happen for the interaction term, because the high effort by a productive worker j should be more likely to be rewarded when γ_{ij} increases, i being the company.

By looking at the results in the Table above, we can conclude that the results support reciprocity, since the coefficients for high effort and the interactions are positive and significant.

5 Discussion and Concluding Remarks

There is a great deal of evidence against selfishness, but also infinite ways not to be selfish. In theoretical work, we need the right instrument to discriminate among the models. Controlled experiments are the right tool to use because they make it possible to test specific predictions in isolation. This is how the literature normally proceeds in the domain of other regarding preferences, when comparing outcome based models with intention based models. However, most of those models have degrees of freedom that make them suitable to account ex post for the evidence. By means of a comparative statics theorem, we provide a test which is as robust as the within-subject analysis of Blanco et al. (2011), but with the advantage that we can test one theory against the other.

To build a comparative statics theorem, we study the reciprocity model in detail. Surprisingly, the predictions we derive are relatively general. For example, for both inequity aversion and reciprocity, they hold for an arbitrary number of workers, which is not unusual in the inequity aversion case but it is definitely unusual for psychological game models. Indeed, we show that our reciprocity results are, in some way, flexible, which is not the case for many applications.

Our main experimental results are largely consistent with the theory of reciprocity.

Nevertheless, this is not a generalized rejection of the inequity aversion model. In many settings and environments, the experimental data are consistent with its predictions, and it has an undeniable advantage in terms of low analytical cost. However, the inequity aversion model loses predictive power when the perfect equality is outside the feasibility set (Nelson, 2002), where agents seem to take intentions into account (for example, individuals do not punish an agent who has tried but failed to implement an equitable outcome). Additionally, Yang, Onderstal, and Schram (2016) conclude that the predictive power of the inequity aversion model is harmed when we open up the possibility to reciprocate.

Finally, we conclude that much work is still needed on psychological game theory (Dufwenberg & Patel, 2019). This class of models is not widely used because of the difficulty of agreeing on how to model higher order beliefs. In spite of this, the evidence shows that intentions matter, and there is a crucial role played by second order beliefs in social norms (Bicchieri, 2017). Since the latter has recently received a lot of attention, we should expect further research on how to model beliefs and on psychological game theory in general. Using this class of models in specific settings, as we have demonstrated in this article, is a way to derive specific predictions and to further contribute to deepen our theoretical understanding of intention based interactions.

References

Akerlof, G. A. (1982). Labor Contracts as Partial Gift Exchange. *The Quarterly Journal of Economics*, *97*(4), 543–569. Retrieved from `http://www.jstor.org/stable/1885099` doi: 10.2307/1885099

Battigalli, P., & Dufwenberg, M. (2009, jan). Dynamic psychological games. *Journal of Economic Theory*, *144*(1), 1–35.

Bicchieri, C. (2017). *Norms in the wild: How to diagnose, measure, and change social norms.* doi: 10.1093/acprof:oso/9780190622046.001.0001

Blanco, M., Engelmann, D., & Normann, H. T. (2011, jun). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior*, *72*(2), 321–338. Retrieved from `https://www.sciencedirect.com/science/article/abs/pii/S089982561000148X` doi: 10.1016/J.GEB.2010.09.008

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

Çelen, B., Schotter, A., & Blanco, M. (2017, may). On blame and reciprocity: Theory and experiments. *Journal of Economic Theory*, *169*, 62–92.

Charness, G., & Haruvy, E. (2002, jan). Altruism, equity, and reciprocity in a gift-exchange experiment: an encompassing approach. *Games and Economic Behavior*, *40*, 203–231.

Charness, G., & Kuhn, P. (2011). *Lab Labor: What Can Labor Economists Learn from the Lab?* doi: 10.1016/S0169-7218(11)00409-6

Dufwenberg, M., & Kirchsteiger, G. (2004, jan). A theory of sequential reciprocity. *Games and Economic Behavior*, *47*, 268–298.

Dufwenberg, M., & Kirchsteiger, G. (2019). Modelling kindness. *Journal of Economic Behavior and Organization*, *167*, 228–234. doi: 10.1016/j.jebo.2018.07.014

Dufwenberg, M., & Patel, A. (2019). Introduction to special issue on psychological game theory. *Journal of Economic Behavior and Organization*, *167*, 181–184. doi: 10.1016/j.jebo.2019.05.034

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, *54*(2), 293–315.

Fehr, E., & Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868. Retrieved from `http://www.jstor.org/stable/2586885`

Fischbacher, U. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*. doi: 10.1007/s10683-006-9159-4

Geanakoplos, J., Pearce, D., & Stacchetti, E. (1989, jan). Psychological games and sequential rationality. *Games and Economic Behavior*, *1*(1), 60–79. doi: 10.1016/0899-8256(89)90005-5

Isoni, A., & Sugden, R. (2019). Reciprocity and the Paradox of Trust in psychological game theory. *Journal of Economic Behavior and Organization*. doi: 10.1016/j.jebo.2018.04.015

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

Nelson, W. R. (2002). Equity or intention: It is the thought that counts. *Journal of Economic Behavior and Organization*, *48*(4), 423–430. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0167268101002451` doi: 10.1016/S0167-2681(01)00245-1

Niederle, M., & Vesterlund, L. (2007). *Do women shy away from competition? Do men compete too much?* doi: 10.1162/qjec.122.3.1067

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

Segal, U., & Sobel, J. (2007). Tit for tat: Foundations of preferences for reciprocity in strategic settings. *Journal of Economic Theory*, *136*(1), 197–216. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0022053106001207` doi: https://doi.org/10.1016/j.jet.2006.07.003

Yang, Y., Onderstal, S., & Schram, A. (2016). Inequity aversion revisited. *Journal of Economic Psychology*, *54*, 1–16. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0167487015001518` doi: https://doi.org/10.1016/j.joep.2015.12.009

A Proofs of reciprocity claims

In this Appendix, we present the formal proofs of the claims and theorem. In first place, take $i \in W$ a worker and j, k any type of agent, then:

- $p_{1i} \equiv \text{Prob}(e^i = e_1)$, $p_{2i} \equiv \text{Prob}(e^i = e_2)$ and $p_{3i} \equiv \text{Prob}(e^i = e_3)$. Also, $\sigma_i \equiv (p_{1i}, p_{2i}, p_{3i})$,
- $e \equiv (e^j)_{j \in W}$. Moreover, $r_i(e) \equiv \text{Prob}(b_i = b|e)$ and $r(e) \equiv ((r_j(e))_{j \in W}, 1 - \sum_{j \in W} r_j(e))$. Observe that $r_i(e)$ is the probability that worker i gets the bonus in the history determined by e. Thus, one manager's behavioral strategy can be written by $r = (r(e))_{e \in E \times E}$, in other words, a behavioral strategy for M is a combination of probability distributions over her actions in each possible history of the game,
- p^j_{ni} is the belief of j over p_{ni}. Likewise is defined $r^j_i(e)$, σ^j_i, $r^j(e)$ and r^j, and
- p^{jk}_{ni} is the belief of k over p^j_{ni}. Likewise we define $r^{jk}_i(e)$, σ^{jk}_i, $r^{jk}(e)$ and r^{jk}.

Now, recalling that the manager chooses upon knowing the effort implemented by each worker, and according to the belief's actualization mechanism, the manager assumes that those efforts were chosen with probability 1. Next, we have to construct $\hat{u}_M$ contingent upon each possible combination of workers' efforts $e \in E \times E$. Thus, given $i \in W$, holds:

$$EV_M[\pi_i](e) = w + \frac{b}{2} - c(e^i). \tag{6}$$

Consequently:

$$\kappa_{Mi}(e, r(e)) = b\left(r_i(e) - \frac{1}{2}\right). \tag{7}$$

Similarly, we have to determine M's belief about i's kindness with M. Since in history h the workers' choices are observed, the equitable payoff of i with M is:

$$EV_i[\pi_M](e, r^i(e)) = e_2 + e^{-i} - \sum_{j \in W} r_j^j(e)b. \tag{8}$$

Therefore:

$$\lambda_{MiM}(e, r^{iM}(e)) = e^i - e_2. \tag{9}$$

Equation (7) implies that the manager considers herself to be kind with the worker i if and only if the probability assigned to give the bonus to him is greater than 0.5. Likewise, Equation (9) implies that the manager considers that the worker has been kind with her if he chooses e_3,[9] since the referential effort is e_2.

From the former equations, we construct $\hat{u}_M$ given the history h as:

$$\hat{u}_M(e, r(e)) = \sum_{j \in W}(e^j - r_j(e)b) + \sum_{j \in W}\left(\gamma_{Mj} b\left(r_j(e) - \frac{1}{2}\right)(e^j - e_2)\right).$$

Which is equivalent to:

$$\hat{u}_M(e, r(e)) = \sum_{j \in W}\left(e^j + r_j(e)b(\gamma_{Mj}(e^j - e_2) - 1) - \frac{1}{2}\gamma_{Mj} b(e^j - e_2)\right). \tag{10}$$

Herein, we are able to formalize the claims relative to the manager's behavior and provide the respective proof.

[9] This result is due to the following two assumptions. The first assumption is DK's equitable payoff definition. Since they have defined it as the mean of the maximum and minimum payoff; nevertheless, they argue that this definition is made for the sake of simplicity and the robustness of the results when we define the equitable payoff as any arbitrary convex combination between maximum and minimum payoff. The second assumption is that $e_2 = (e_3 + e_2)/2$, because this makes e_2 to be the reference point for the worker's kindness, however, taking e_2 different to $(e_3 + e_2)/2$ changes the manager's perception of the e_2 kindness, then e_2 would be comparable in terms of kindness with another effort level. Thus, taking e_2 as an action that is not kind or unkind is a more interesting case to study.

Claim 1. *Let be $i \in W$ and $e \in E \times E$. If $e^i \neq e_3$, then the manager chooses $r_i(e) = 0$. In other words, when a worker does not choose the highest effort, he will not get the bonus.*

Proof. This behavior is due to Equation (10), specifically the coefficient that multiplies $r_i(e)$ is $b(\gamma_{Mi}(e^i - e_2) - 1)$. If $e^i = e_1$ or $e^i = e_2$, such coefficient is negative. Therefore, the choice that maximizes the manager's psychological utility is $r_i(e) = 0$.

$\square$

Claim 2. *Let be $i \in W$ and $e \in E \times E$. In equilibrium, if $e^i = e_3$, and:*

1. *$e^{-i} \neq e_3$ and $\gamma_{Mi} > \frac{1}{e_3 - e_2}$, the manager chooses $r_i(e) = 1$. In other words, if a worker chooses the highest effort, his co-worker does not do it and the manager is sensitive enough to reciprocity, the worker with the best performance gets the bonus,*
2. *$\gamma_{Mi} < \frac{1}{e_3 - e_2}$, then the manager chooses $r_i(e) = 0$. In other words, even if the worker has chosen the highest effort, the manager does not give the bonus to him when she is not sensitive enough to reciprocity, and*
3. *$e^{-i} = e_3$ and $\gamma_{M-i}, \gamma_{Mi} > \frac{1}{e_3 - e_2}$, the manager gives the bonus to the worker for whom she feels a greater sensitivity to reciprocity.*

Proof. Since the coefficient that multiplies $r_i(e)$ in the equation (10) is $b(\gamma_{Mi}(e_3 - e_2) - 1)$. Then:

1. by claim 1, $r_{-i}(e) = 0$. Now, such a coefficient is positive if and only if $\gamma_{Mi} > \frac{1}{e_3 - e_2}$. Consequently, $r_i(e) = 1$,
2. since $\gamma_{Mi} < \frac{1}{e_3 - e_2}$, the coefficient is negative, which implies that the manager's optimal behavior is $r_i(e) = 0$, and
3. given item 1 of this proof, for every $k \in W$ the coefficients that multiplies $r_k(e)$ are positive. Nevertheless, $\gamma_{M-i} > \gamma_{Mi}$ entails $b(\gamma_{M-i}(e_3 - e_2) - 1) > b(\gamma_{Mi}(e_3 - e_2) - 1)$. Thus, assigning a higher probability to give the bonus to $-i$ increases psychological utility for the manager. Therefore, the manager gives the bonus to whom she feels a greater sensitivity to reciprocity.

$\square$

On the other hand, in the workers' analysis we have to highlight two issues. The first is that workers choose their effort simultaneously, and before the manager, which implies that they make a more complex decision. Thus, we take advantage of the manager's optimal behavior to characterize the workers' behavior since they are able to anticipate it. The second issue is that in the specific context analyzed, there is no causal role for kindness between workers. To see this, suppose that the manager is more sensitive to reciprocity towards i rather than his co-worker, $-i$. Thus, $-i$'s action does not affect the likelihood of bonus for i, since such probability only depends on i's effort.[10] Therefore, $-i$'s action does not have any impact on i's material payoff.

Claim 3. *Let be $i \in W$. In equilibrium, if $r_i^i = 0$, then i chooses $p_{1i} = 1$. In other words, if a worker believes that the manager will not give the bonus to him regardless of his choice, he*

[10]Technically, reciprocity between workers may matter when the manager is equally sensitive to both, but this is an event with zero probability.

implements the lowest effort. In particular, whether $\gamma_{Mi} < \frac{1}{e_3-e_2}$ *(Claim 2.2.) or* $\gamma_{Mi} < \gamma_{M-i}$ *and* $p^i_{3-i} = 1$ *(Claim 2.3).*

Proof. We have to construct $\hat{u}_i$ under the belief that the manager will not give the bonus to him:

$$EV_i[\pi_M](\sigma^i_{-i}, r^i) = e_2 + \sum_{n=1}^{3} p^i_{n-i} e_n - b p^i_{3-i} r^i_{-i}$$

$$\kappa_{iM}(\sigma_i, \sigma^i_{-i}, r^i) = \sum_{n=1}^{3} p_{ni} e_n - e_2. \tag{11}$$

It should be noted that Equation (11) can be interpreted as follows: a worker thinks that he has been kind to the manager if the expected effort of his strategy is large enough with respect to the medium effort. Now:

$$EV_M[\pi_i]((\sigma^M_j)_{j\in W}) = w + \frac{b}{2} - \sum_{n=1}^{3} p^M_{ni} c(e_n).$$

Then:

$$\lambda_{iMi}((\sigma^{Mi}_j)_{j\in W}, r^i) = b\left(p^{Mi}_{3i} r^i_i - \frac{1}{2}\right). \tag{12}$$

Taking into account that the manager chooses after she has observed the workers' choices, and therefore, the manager's belief is actually the action implemented by the worker (and the worker is aware about it), we have that $\sigma^{Mi}_j = \sigma^i_j$ for all $j \in W$. In particular, $p^{Mi}_{3i} = p^i_{3i} = p_{3i}$. Moreover, since the worker believes that he will not get the bonus, then the formula for the kindness becomes:

$$\lambda_{iMi}((\sigma^{Mi}_j)_{j\in W}, r^i) = -\frac{b}{2}. \tag{13}$$

Thus, Equation (13) shows that under this assumption, the worker considers the manager to be unkind. However, in this case $\hat{u}_i$ is:

$$\hat{u}_i(\sigma_i, \sigma^i_{-i}, r^i) = w - \sum_{n=1}^{3} p_{ni} c(e_n) - \frac{\gamma_{iM} b}{2}\left(\sum_{n=1}^{3} p_{ni} e_n - e_2\right).$$

Notice that positive p_{ni} reduces both the material payoffs and the psychological payoffs of the worker. Thus, he must choose the p_{ni} that generates the lowest effort; in other words, he chooses $p_{1i} = 1$.

□

Claim 4. *Let* $i \in W$. *If* $r^i_i = 1$, *then* $p_{3i} = 1$. *This means that if a worker believes that he will get the bonus when he chooses the highest effort, then he will act accordingly. In particular, it will happen when* $\gamma_{Mi} > \frac{1}{e_3-e_2}$ *and* $\gamma_{Mi} > \gamma_{M-i}$ *(Claim 2.3.) or* $\gamma_{Mi} > \frac{1}{e_3-e_2}$ *and* $p^i_{-i3} = 0$ *(Claim 2.1.).*

Proof. Given these beliefs, we have:

$$EV_i[\pi_M](\sigma^i_{-i}, r^i) = e_2 + \sum_{n=1}^{3} p^i_{n-i} e_n - b\left(\frac{1 + p^i_{3-i} r^i_{-i}}{2}\right)$$

$$\kappa_{iM}(\sigma_i, \sigma^i_{-i}, r^i) = \sum_{n=1}^{3} p_{ni} e_n - e_2 + b\left(\frac{1 + p^i_{3-i} r^i_{-i}}{2} - p_{3i} - (1 - p_{3i}) p^i_{3-i} r^i_{-i}\right). \qquad (14)$$

On the other hand:

$$EV_M[\pi_i]((\sigma^M_j)_{j\in W}) = w + \frac{b}{2} - \sum_{n=1}^{3} p^M_{ni} c(e_n).$$

Therefore:

$$\lambda_{iMi}((\sigma^{Mi}_j)_{j\in W}, r^i) = b\left(p^{Mi}_{3i} - \frac{1}{2}\right) = b\left(p_{3i} - \frac{1}{2}\right) \qquad (15)$$

Observe that Equation (15) shows that a worker thinks that the manager has been kind with him if she gives him the bonus with a probability greater than $\frac{1}{2}$, which is the same probability of implementing the highest effort (given the hypothesis). Thus, the psychological utility is:

$$\begin{aligned}\hat{u}_i(\sigma_i, \sigma^i_{-i}, r^i) &= w + p_{3i} b - \sum_{n=1}^{3} p_{ni} c(e_n) \\ &+ b\gamma_{iM}\left(p_{3i} - \frac{1}{2}\right)\left(\sum_{n=1}^{3} p_{ni} e_n - e_2 + b\left(\frac{1 + p^i_{3-i} r^i_{-i}}{2} - p_{3i} - (1 - p_{3i}) p^i_{3-i} r^i_{-i}\right)\right).\end{aligned}$$

Notice that:

1. $\hat{u}_i(e_3, \sigma^i_{-i}, r^i) - \hat{u}_i(e_2, \sigma^i_{-i}, r^i) = b - [c(e_3) - c(e_2)] + \frac{b\gamma_{iM}}{2}(e_3 - e_2) > 0$ because of Assumption 1,[11] and
2. $\hat{u}_i(e_3, \sigma^i_{-i}, r^i) - \hat{u}_i(e_1, \sigma^i_{-i}, r^i) = b - [c(e_3) - c(e_1)] > 0$, also by Assumption 1.

Thus, independently of his beliefs and γ_{iM}, the worker always chooses the highest effort. □

Finally, we weaken the standard assumption that the sensitivity parameters are common knowledge and assume instead that there exists a joint probability density function $f(\gamma_{Mi}, \gamma_{M-i})$ defined over $\mathbb{R}^2_+$, and we denote by $F(\gamma_{Mi}, \gamma_{M-i})$ the associated joint cumulative distribution function.

If P_i represents the bonus probability of the worker i when he has chosen the highest effort, then $P_i = \int_0^\infty \int_{\gamma_{M-i}}^\infty f(\gamma_{Mi}, \gamma_{M-i}) d\gamma_{Mi} d\gamma_{M-i} - \frac{1}{2(e_3-e_2)^2}$. Thus, the problem of the worker i is:

$$\max_{\sigma_i} E_{(\gamma_{Mi}, \gamma_{M-i})} \hat{u}_i,$$

which is equivalent to:

$$\max_{\sigma_i} P_i(\hat{u}_i | r^i_i = 1) + (1 - P_i)(\hat{u}_i | r^i_i = 0).$$

Claim 5. *Worker i implements the highest effort if and only if P_i is large enough. Otherwise, he chooses the lowest effort.*

[11] Specifically, $b > c(e_3) - c(e_1) > c(e_3) - c(e_2)$.

Proof. For the sake of simplicity, we denote by $E_{(\gamma_{Mi},\gamma_{M-i})}\hat{u}_i \equiv E\hat{u}_i$. Now, let us see that $E[\hat{u}_i(e_1)] > E[\hat{u}_i(e_2)]$, which implies that worker i never chooses the medium effort. Indeed, $E[\hat{u}_i(e_1)] - E[\hat{u}_i(e_2)] = c(e_2) - c(e_1) + \frac{b\gamma_{iM}}{2}(e_2 - e_1) > 0$.

On the other hand, worker i chooses the highest effort if his expected psychological utility is higher than the expected psychological utility of the lowest effort, formally, if $E[\hat{u}_i(e_3)] - E[\hat{u}_i(e_1)] > 0$. Taking into account that:

$$\begin{aligned} E[\hat{u}_i(e_3)] - E[\hat{u}_i(e_1)] &= P_i\,[b + c(e_1) - c(e_3)] \\ &\quad + (1 - P_i)\left[c(e_1) - c(e_3) - \frac{b\gamma_{iM}}{2}(e_3 - e_1)\right] \\ &= P_i\left[b + \frac{b\gamma_{iM}}{2}(e_3 - e_1)\right] + c(e_1) - c(e_3) \\ &\quad + \frac{b\gamma_{iM}}{2}(e_3 - e_2), \end{aligned}$$

then, $E[\hat{u}_i(e_3)] - E[\hat{u}_i(e_1)] > 0$ if and only if $P_i > \frac{c(e_3)-c(e_1)+\frac{b\gamma_{iM}}{2}(e_3-e_1)}{b+\frac{b\gamma_{iM}}{2}(e_3-e_1)}$. Finally, it is important to say that the previous threshold belongs the (0,1) interval, because of Assumption 1. □

B Proofs of inequity aversion claims

Claim 6. *In equilibrium, each worker implements the lowest effort.*

Proof. To start, suppose that worker i believes that he will not get the bonus, then the material payoff of his co-worker, $-i$, will be at least as large as his own. Thus:

$$\tilde{u}_i(\pi) = w - c(e^i) - \frac{\alpha_i}{2}(e^i + e^{-i} - b_{-i} + w - c(e^{-i}) + b_{-i} - 2w + 2c(e^i))$$

Observe that an increase in e^i reduces utility, whereby the worker i chooses the lowest effort.

Now suppose that worker i believes that he will get the bonus. In this case, he obtains a greater material payoff than his co-worker, which makes him feel guilty. Formally:

$$\tilde{u}_i(\pi) = w + b - c(e^i) - \frac{\alpha_i}{2}(e^i + e^{-i} - b - w - b + c(e^i)) - \frac{\beta_i}{2}(w + b - c(e^i) - w + c(e^{-i})).$$

From this utility function, the part involved with the worker's choice variable is $c(e^i)\left(\frac{\beta_i-\alpha_i}{2} - 1\right) - \frac{\alpha_i}{2}e^i$.

Thus, his choice will be $e^i = e_1$ if:

- $c(e_1)\left(\frac{\beta_i-\alpha_i}{2} - 1\right) - \frac{\alpha_i}{2}e_1 > c(e_3)\left(\frac{\beta_i-\alpha_i}{2} - 1\right) - \frac{\alpha_i}{2}e_3$. Manipulating this expression, we have $\frac{e_3-e_1}{c(e_3)-c(e_1)} > \frac{\beta_i-\alpha_i-2}{\alpha_i}$. Taking into account that the Assumption 1 implies that $\frac{c(e_3)-c(e_1)}{e_3-e_1} < \frac{1}{2}$, which is equivalent to $\frac{e_3-e_1}{c(e_3)-c(e_1)} > 2$, it follows that a sufficient condition for the choice of the lowest effort over the highest one corresponds to $\frac{\beta_i-\alpha_i-2}{\alpha_i} < 2$. By algebraic manipulation, it is possible to write the previous inequality as $\beta_i - 2 < \alpha_i$; that is, worker i chooses the lowest effort if he is envious enough.

Finally, bearing in mind that $\beta_i < 1$, it holds that $\beta_i - 2 < 0 \leq \alpha_i$. This allow us to conclude that worker i prefers the lowest effort over the highest one when he gets the bonus independently of his envy parameter, since in this case, being envious enough means $\alpha_i \geq 0$, and

- $c(e_1)\left(\frac{\beta_i-\alpha_i}{2} - 1\right) - \frac{\alpha_i}{2}e_1 > c(e_2)\left(\frac{\beta_i-\alpha_i}{2} - 1\right) - \frac{\alpha_i}{2}e_2$. Manipulating this expression, it follows that $\frac{e_2-e_1}{c(e_2)-c(e_1)} > \frac{\beta_i-\alpha_i-2}{\alpha_i}$. By strict convexity of the cost function and by $e_2 = \frac{e_3+e_1}{2}$, it holds $c(e_2) = c\left(\frac{e_3+e_1}{2}\right) < \frac{c(e_3)+c(e_1)}{2}$. Consequently, $\frac{e_2-e_1}{c(e_2)-c(e_1)} > \frac{e_3-e_1}{c(e_3)-c(e_1)}$. Thus, a sufficient condition for the choice of the lowest effort over the medium one is $\frac{e_3-e_1}{c(e_3)-c(e_1)} > \frac{\beta_i-\alpha_i-2}{\alpha_i}$, which according to the previous item, holds in every case.

□

Claim 7. *The manager gives the bonus to any worker if and only if her inequity aversion is large enough.*

Proof. The manager's utility function is:

$$
\begin{aligned}
\tilde{u}_M(\pi) &= \pi_M - \frac{\beta_M}{2}\sum_{j\in W}\max\{\pi_M - \pi_j, 0\} \\
&= e^i + e^{-i} - b_i - b_{-i} - \frac{\beta_M}{2}(2e^i + 2e^{-i} - 3b_i - 3b_{-i} - 2w + c(e^i) + c(e^{-i})) \\
&= (-b_i - b_{-i})\left(1 - \frac{3\beta_M}{2}\right) + e^i + e^{-i} - \frac{\beta_M}{2}(2e^i + 2e^{-i} - 2w + c(e^i) + c(e^{-i})).
\end{aligned}
$$

Taking into account that the manager chooses b_i, b_{-i}, it is possible to conclude that she gives the bonus to any worker if $\beta_M > \frac{2}{3}$; in other words, if she is averse enough to guilt.

Finally, observe that the worker's behavior does not affect the manager's decision. This entails that the manager's behavior does not change under incomplete information. The same holds in the worker's case. □

C Proof of the main theorem

Theorem 1. *Take $j \in W$ and e_{nj} the value of the effort e_n for worker j. If e_{3i} increases (i.e., worker i suffers a productivity shock) and the agents are motivated by reciprocity, then the probability of getting the bonus is increasing, conditional to i's high effort choice. Also, the probability that worker i implements the high effort increases by the productivity shock. In the case where the agents are motivated by inequity aversion, these probabilities are invariant.*

Proof. 1. Since $e_{3i} > e_{3-i}$ and $e_{3j} - e_{2j} = \frac{e_{3j}-e_{1j}}{2}$ for every $j \in W$, we have:

$$P_i = \int_0^\infty \int_{\gamma_{M-i}}^\infty f(\gamma_{Mi}, \gamma_{M-i})d\gamma_{Mi}d\gamma_{M-i} - 2(e_{3i} - e_{1i})^{-2}.$$

Therefore:

$$\frac{\partial P_i}{\partial e_{3i}} = 4(e_{3i} - e_{1i}) > 0,$$

this implies that for a worker, the likelihood to get the bonus upon choosing the high effort is increasing in e_3. Thus, a productivity shock has a positive impact on the probability of a bonus, conditional on the high effort choice.

On the other hand, if we denote by $T(e_3) = \frac{c(e_3)-c(e_1)+\frac{b\gamma_{iM}}{2}(e_3-e_1)}{b+\frac{b\gamma_{iM}}{2}(e_3-e_1)}$ (the probability threshold which guarantee the high effort choice), then:

$$T'(e_3) = \frac{\frac{b\gamma_{iM}}{2}\left(b+\frac{b\gamma_{iM}}{2}(e_3-e_1)\right) - \left(c(e_3)-c(e_1)+\frac{b\gamma_{iM}}{2}(e_3-e_1)\right)\frac{b\gamma_{iM}}{2}}{\left(b+\frac{b\gamma_{iM}}{2}(e_3-e_1)\right)^2},$$

since $c'(e_3) = 0$ (the effort cost remains the same by effort level, independently of the productivity) and $b - c(e_3) > c(e_1)$ (Assumption 1), it follows that the numerator is negative. Thus, $T'(e_3) < 0$. Consequently, this shock makes the choice of the highest effort more likely.

2. Remember that an inequity averse worker always chooses the lowest effort. Since the cost function remains the same, the productivity shock does not affect the workers' material payoffs. For this reason, the manager's behavior is invariant to the shock.

□

D Extensions

In this Appendix, we prove that the comparative statics holds for any norm of fairness, allowing for heterogeneity among agents.

Consider agent i's personal norm of fairness towards j defined by:

$$EV_i[\hat{\pi}_j]((b_{ij})_{j\neq i}) \equiv \alpha_{ij} \max_{a_i\in A_i} \hat{\pi}_j(a_i,(b_{ij})_{j\neq i}) + (1-\alpha_{ij}) \min_{a_i\in A_i} \hat{\pi}_j(a_i,(b_{ij})_{j\neq i}), \tag{16}$$

where $\alpha_{ij} \in (0,1)$. The belief of i regarding the degree of kindness by j towards himself is defined by:

$$\lambda_{iji}(b_{ij}(h),(c_{ijk}(h))_{k\neq j}) \equiv \hat{\pi}_i(b_{ij}(h),(c_{ijk}(h))_{k\neq j}) - EV_j[\hat{\pi}_i]((c_{ijk}(h))_{k\neq j})). \tag{17}$$

As a consequence, the reference point used by the manager to classify i's effort as kind or unkind is $e_i^* \equiv \alpha_{iM}e_3 + (1-\alpha_{iM})e_1 \in (e_1, e_3)$. Thus, e_3 and e_1 are still a kind and unkind strategy for each worker, respectively. Nevertheless, e_2 could be a kind or unkind action, and for that reason, the full characterization of the equilibria is slightly more complex. However, from the manager's utility function (Equation 10), it is still true that she rewards i's high effort if $\gamma_{Mi} > \frac{1}{e_3-e_i^*}$, and in case both workers choose the high effort, she rewards i against $-i$ if her sensitivity to reciprocity towards the worker weighted by the kindness degree of his action is higher for i than for $-i$; formally, if $\gamma_{Mi}(e_3-e_i^*) > \gamma_{M-i}(e_3-e_{-i}^*)$.

Observe that if $e_3 - e_i^*$ gets larger, both inequalities are satisfied with a lower sensitivity to reciprocity γ_{Mi}. Consequently, given the joint probability distribution $f(\gamma_{Mi},\gamma_{M-i})$, it is more likely to get the bonus when choosing the high effort.[12]

[12] Technically, we are expanding the set where worker i gets the bonus conditional upon the choice of the high effort, and since all probability measure is increasing, it generates a higher probability of being rewarded.

Notice that, by algebraic manipulation, $e_3 - e_i^* = (e_3 - e_1)(1 - \alpha_{iM})$, which clarifies the effect of an increase in e_3 over $e_3 - e_i^*$. Finally, we can conclude that our main theorem holds when each agent has its own norm of fairness.