Peer Effects, Reciprocity and Equilibrium Preferences

Jose A. Carrasco* Rodrigo Harrison† Mauricio G. Villena‡

July 4, 2021

Abstract

We model how individual preferences are shaped by peer effects. Our model also accounts for heterogeneous players — with altruistic, selfish or spiteful preference —who randomly engage in pairwise interactions. To explore social influences, we allow them to act reciprocally to adjust their preferences depending on with whom they interact. How much their preference changes is our measure of peer effects. We show that peer elects always arise in equilibrium, but intensity varies among players. In particular, we find that peer effects are reduced for extreme types: for players who are too altruistic or too spiteful. Otherwise, we might observe preference-reversion: due to large peer effects, an intrinsically altruistic (spiteful) player expects to behave spitefully (altruistically). Our model also predicts that in frameworks characterized by positive externalities and strategic complements (substitutes), reciprocity choices become strategic substitutes (complements), and peer effects are larger. A stochastically better opponent's type distribution also leads to larger peer effects and more expected altruistic behavior. We also show that when player's types are unknown, there is no preference-eversion, and equilibrium preferences are as selfish as possible.

JEL codes: C72, A13

Keywords: Peer Effects, Reciprocity, Endogenous Preferences, Incomplete Information

*Corresponding Author. Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibañez, 2640 Diagonal Las Torres, Santiago, Chile, email: jose.carrasco@uai.cl, web: www.tonocarrasco.com

†Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibañez, 2640 Diagonal Las Torres, Santiago, Chile, email: rodrigo.harrison@uai.cl, web: www.rodrigoharrison.com

‡Facultad de Economia y Empresa, Universidad Diego Portales, Avenida Santa Clara 797, Huechuraba, Santiago, Chile, email: mauricio.villena@udp.cl, Web: www.mauriciovillena.com

1 Introduction

People sometimes adjust their social preferences and behavior depending on what they see other individuals do. As such, we observe agents exhibiting positive concern towards others or behaving selfishly towards others as being influenced by what their peers do. Indeed, although cooperation and altruism lie at the heart of human lives, there are other situations in which we lash out to do harm, which are also influenced by others. Recognizing these regularities in human actions, the economics literature has extensively studied *peer effects*, which refers to a situation in which an agent is influenced in his or her actions by what a comparison agent does, even if there are no material exchanges between those individuals.[1] Nevertheless, while preferences and peer effects are well documented in the empirical literature, little is known about how peers effects shape social preferences (Thöni and Gächter, 2015). There is no theoretical work squarely dealing with the formal analysis of the link between peer effects and preferences to the best of our knowledge. In this context, this work attempts to contribute by proposing a model to describe how people adjust their preferences, particularly the concern they show for others, influenced by peer effects.

Unlike most previous works in which preferences are modeled as fixed (see for instance theories of inequality aversion people's social preferences are modeled as individually fixed distastes for inequitable outcomes) to provide a theory linking peer effects and preferences, fixed preference assumptions must be relaxed. To wit, our model's key ingredient is that preferences are subject to peer effects, influenced by the behavior of others. As people interact and are constantly exposed to others, their preferences are not immune to their peer influences and therefore they might change.

Our modeling approach accounts for strategic interaction at the preference level and alternative information structures. Specifically, our model accounts for pairwise random meetings Okuno-Fujiwara and Postlewaite (1995) between heterogeneous players that inherit an *intrinsic preference* — an altruistic, selfish or spiteful preference — that randomly engages in a simultaneous-move short-run game (e.g., Cournot, Bertrand). Crucially, players also engage in long-run strategic interaction at the preference level, where peer effects arise: they play an underlying reciprocity game. That is, reciprocity is our model's key ingredient for endogenous *induced preferences* and peer effects. These equilibrium preferences are match-specific and are shaped by how peers reciprocate. How induced preferences differ from their intrinsic values is our measure of peer effects.

[1] See for instance:Chen et al. (2010), Falk et al. (2013), Falk and Ichino (2006a), Frey and Meier (2004), Gächter and Thöni (2010), Gächter et al. (2012), Gächter et al. (2013), Ho and Su (2009),Kremer and Levy (2008), Mas and Moretti (2009), Mittone and Ploner (2011) and Sacerdote (2001).

Typically, previous theoretical works on preferences mainly predict that there are no peer effects, and positive concern for others (altruism) arises only when the strategic context is one of the strategic complements; otherwise, negative concern (spitefulness) arises (see, for instance, Bester and Güth (1998), Bolle (2000), Possajennikov (2000) and Carrasco et al. (2018)). Unlike them, we find that there are peer effects and that players' preferences are altruistic, selfish, or spiteful, dependent on the strategic environment and information structures. They depend on players' own intrinsic types, the information they use to infer their opponent's type, whether the opponent's type is common knowledge, the type of game players engage in, and whether the short-run game is one of the strategic complements or strategic substitutes. This allows us to explain people's actions without assuming that intrinsic preferences change; what changes are the peers, who influence the preferences and behaviors of the individual.

We formally present the model in Section 2, and in Section 3, we fully characterize equilibrium reciprocity and induced preferences when the information structure is such that players' types are common knowledge. To account for different strategic environments, we introduce a parameter $-1 \leq k \leq 1$, where $|k|$ captures the degree of strategic interaction in the short-run game. We refer to the environment as one of the negative externalities and strategic substitutes when $k < 0$ and one of the positive externalities and strategic complements when $k > 0$. Regardless, we show that peer effects exist, but not necessarily to the same extent among players (Proposition 1), and that this is true regardless of the information structure of the game considered.

In these cases, players can choose a type-specific reciprocity strategy that is not only type-dependent, on both their own and their opponent's intrinsic type, but that also depends on the strategic environment. We show that when the short-run game is one of the strategic substitutes with $k < 0$ (complements with $k > 0$), the reciprocity game becomes one of the strategic complements (substitutes). Moreover, as k rises, the more intrinsically altruistic player grows less reciprocal, whereas the other grows more reciprocal. That is, reciprocity's best responses might be downward- or upward-sloping functions, depending on the strategic context. As a result, altruistic, spiteful, or selfish preferences and their intensity arise crucially and depends on how players' types compare to the degree of strategic interaction k.

We then measure peer effects by exploring how much preferences differ from their intrinsic value. In other words, do players expect to behave more altruistically or more spitefully than what they intrinsically are? Furthermore, is it possible for an intrinsically altruistic (spiteful) player to expect to behave spitefully (altruistically)? We compute the expected interim preferences, and we show that a stochastically better opponent's type distribution, as well as larger values of k, leads to more altruistic expected behavior. Such behavior also depends on a player's own type: both sufficiently altruistic and spiteful players expect to reduce their intensity of concern for others. Moreover, for those players that are neither sufficiently altruistic nor sufficiently spiteful, we might observe preference-reversion, the fact that in expectation, an intrinsically altruistic (spiteful) player behaves spitefully (altruistically). In these cases, induced preferences might reverse depending on the strategic context (Proposition 2).

In Section 4, we explore contexts that are better characterized by an information structure such that players' types are unknown (i.e., the incomplete information case). We find that the optimal reciprocity choice is a dominant strategy, and players reciprocate by weighting their opponent's expected type. Interestingly, and unlike the perfect information case, optimal reciprocity is independent not only of the opponent's type, as one might expect, but also of the strategic environment summarized in parameter k. In addition, contrary to the perfect information case, optimal reciprocity yields interim expected preferences that are as selfish as they can be. We find that a necessary and sufficient condition for altruism (spite) to arise is either intrinsic altruism (spitefulness) by one player or that, on average, players are altruists (spiteful) (Proposition 3). We also show (see Proposition 4) that under a context of incomplete information, in which we assume that the specific type of the opponent is unknown and that each player only knows the distribution, *preference-reversion* is not possible. This indicates that even in the presence of large peer effects, an intrinsically altruistic (spiteful) player will never change towards spiteful (altruistic) behavior. These results seem to suggest that the effects of social influence (peer effects) on individual's behavior are somehow diluted under a context of incomplete information on other player types.

LITERATURE: Our framework is related to the literature on interdependent preferences and on the literature on peer effects. Unlike our work, the interdependent preferences approach (see, *inter alia*; Sobel (2005), Fehr and Schmidt (1999), Güth and Napel (2006), Charness and Rabin (2002), Koçkesen et al. (2000), Alger and Weibull (2013) and Isoni and Sugden (2019)) typically considers exogenously specified contexts or fixed preferences that are not influenced by others' behavior. More recently, Carrasco et al. (2018) explore the evolutionary stability of interdependent preferences in a context with perfect information and a strategic environment that shows negative externalities and strategic substitutes. We depart from these previous works because our framework is not evolutionary or cultural transmission based, and it considers players' optimizing behavior and strategic interaction under a rich set of different contexts.

The literature on peer effects is large but has focused mostly on the field of education.

The notion that peer effects are important to educational outcomes has been confirmed both theoretically (Arnott and Rowse, 1987; Benabou, 1993; Lazear, 2001; McMillan, 2004) and empirically (Sacerdote, 2001; Zenou et al., 2014). Another branch of the empirical literature on peer effects that focuses on other economically important settings also confirms that people's behavior is often shaped by others' behavior. Some examples include retirement savings decisions (Duflo and Saez (2002), Beshears et al. (2015)) corruption (Dong et al. (2012)), drug and alcohol consumption Kremer and Levy (2008); Gaviria and Raphael (2001) and behavior at work Bandiera et al. (2010); Guryan et al. (2009); Waldinger (2011); Ichino and Maggi (2000); Mas and Moretti (2009).

Interdependent preferences and peer effects are well documented, and although they are highly connected topics, there is still a large gap between them. The experimental work by Gächter et al. (2013) suggests that social preference models (instead of the social norm approach) provide a "parsimonious explanation" for peer effects. From a theoretical viewpoint, an exception is a recent work by Fershtman and Segal (2018), which we see as an attempt to connect preferences and social influence. The authors explore the properties of social influence functions and their effect on equilibrium behavior. Unlike us, Fershtman and Segal (2018) do not account for strategic behavior at the preference level. Even if players are aware that they influence others, they do not behave strategically, which is a distinct property of our model.

In addition, our paper relates to the experimental work that has tested peer effects(Zimmerman, 2003; Falk and Ichino, 2006b; Gächter and Thöni, 2010; Bougheas et al., 2013). In a relatively recent contribution, Thöni and Gächter (2015) present an experimental gift-exchange game with effort revision to study the role of peer effects in social preferences. They find that efforts are strategic complements and that theories of reciprocity do not predict peer effects. Our model provides a tractable approach that analyzes peer effects and preferences in different strategic contexts: strategic complements or strategic substitutes. In another experiment on peer effects, Bardsley and Sausgruber (2005) finds that reciprocity accounts for roughly two-thirds of the "crowding-in" tendency in public goods provision.

This paper is organized as follows. We present the model in Section 2 and offer theoretical predictions for reciprocity, preferences and peer effects in Section 3. We then explore peer effects when players' types are not commonly known in Section 4. Finally, we present our conclusions in Section 5. All proofs are deferred to the appendix.

2 The Model

We consider two groups, indexed by $i, j \in \{1,2\}$ with $i \neq j$, each with a continuum of players. Meetings happen continuously between players of different groups. Matched players independently choose $x_i \in \mathbb{R}_+$ and $x_j \in \mathbb{R}_+$ (e.g. prices, quantities, etc), respectively. They derive *material payoffs* $\pi_i(x_i, x_j) = x_i(1 - x_i + kx_j)$, where $|k|$ captures the *degree of strategic interaction.* However, preferences are *interdependent*, where β_{ij} summarizes player i concern over his opponents — player j — material payoff; that is, player i perceives *utility* $u_i(x_i, x_j) = \pi_i(x_i, x_j) + \beta_{ij}\pi_j(x_j, x_i)$.[2] To guarantee $x_i, x_j \geq 0$ we let $-1 \leq k \leq 1$, that is the environment allows for both negative externalities and strategic substitutes ($k < 0$) or positive externalities and strategic complements ($k > 0$). This pairwise amount interaction defines a *short run normal form game* $G(\beta, k) = \{\{i,j\}, (x_i, x_j) \in \mathbb{R}^2_+, (u_i, u_j)\}$, whose equilibrium is described by the profile (x_i^*, x_j^*).

We aim to understand interdependent preferences and so our model accounts for two key ingredients. First, for player heterogeneity:[3] within each group players vary by their *type* $\theta \sim F_i$ with continuous densities $f_i \equiv F_i'$ on $\Theta = [-1,1]$ and with $\mathbb{E}_{\theta_i}[\theta_i] = \bar{\theta}_i \in \Theta$.[4] Each player learns his own type previous to the match. Second, for *peer effects* and *long run* strategic interaction at the preference level. We combine these ingredients adopting Levine (1998) specific functional form for preferences letting $\beta_{ij} \equiv \theta_i + \lambda_i(\theta_j - \theta_i)$, where $0 \leq \lambda_i \leq 1$ is a *reciprocity strategy* that weights players *intrinsic preferences* (i.e. types), an invariable component acquired through genetic inheritance. Since $\theta_i, \theta_j \in [-1,1]$, preferences obey $\beta_{ij} \in [-1,1]$. More specifically, given a type profile (θ_i, θ_j), preferences obey $\beta_{ij} \in [\min(\theta_i, \theta_j), \max(\theta_i, \theta_j)]$. We refer to player i as *intrinsically altruistic, selfish or spiteful* if $\theta_i > 0$, $\theta_i = 0$ or $\theta_i < 0$, respectively.

Equilibrium reciprocity — and induced preferences— arise as players exclusively pursue their *long run material payoff* $\pi_i(x_i^*, x_j^*) = \Pi_i$ by choosing their reciprocity coefficient λ_i. This pairwise interaction defines an underlying normal form *reciprocity (long run) game* $\Lambda(k) = \{\{i,j\}, (\lambda_i, \lambda_j) \in [0,1]^2, (\Pi_i, \Pi_j)\}$, whose equilibrium determines how much preferences differ from their intrinsic values and so the size of the peer effects. We say that *there are peer effects for some i if* $\lambda_i > 0$, when in equilibrium player's intrinsic social preference change as a response to other people intrinsic preferences and so $\beta_{ij} \neq \theta_i$.[5]

---

[2] See for instance:Dufwenberg and Kirchsteiger (2019)

[3] Note that if $F_i = F_j$ then both players are drawn from a unique large group.

[4] We use standard notation where $\mathbb{E}_Y[\cdot]$ denotes the expectation in the random variable Y.

[5] In the education literature, there are peer effects if "For given educational resources provided to student A, if having student B as a classmate or schoolmate affects the educational outcome of A"

Later, in Section 4 we relax the information structure of the game and explore equilibrium reciprocity when types are not common knowledge.

3 Peer Effects Under Complete Information

We first solve for equilibrium reciprocity when matched players types are common knowledge. That is, before the interaction each player not only knows his own type, but also his opponent's. We proceed in two steps; first solving the short run game $G(\beta, k)$ and then solving the reciprocity game $\Lambda(k)$, were peer effects arise. Crucially, in this case players are able choose a type specific reciprocity strategy $\lambda_i(\theta_i, \theta_j)$, and so they might reciprocate based not only on their own type, but also on their opponents . In other words, there might be peer effects if $\lambda_i(\theta_i, \theta_j) > 0$.

SHORT RUN GAME: Individuals care not only about their own material payoffs but also about the material payoffs of others.[6] In a meeting they maximize $u_i(x_i, x_j)$, strictly concave in x_i, so that the FOC are necessary and sufficient for a maximum. Best responses are $x_i(x_j) = (1 + kx_j(1 + \beta_{ij}))/2$. As $k, \beta_{ij}, \beta_{ji} \in [-1, 1]$ they each have slopes in $[-1, 1]$. Two extreme cases arise: First, if $\beta_{ij} = \beta_{ji} = 1$ and $k = -1$, then $x_i^* + x_j^* = 1/2$, as best responses perfectly overlap. Second, if $\beta_{ij} = \beta_{ji} = k = 1$ then both players best responses grow linearly without intersecting. Otherwise, if $-1 < k < 1$ or $\min(\beta_{ij}, \beta_{ji}) < 1$, then the unique equilibrium is:

$$x_i^* = \frac{2 + k(1 + \beta_{ij})}{4 - k^2(1 + \beta_{ij})(1 + \beta_{ji})} \tag{1}$$

It follows that x_i^* rises in β_{ji}, so more concern by individual j over player i's material payoff increases x_i^*. Also, that $x_i^* - x_j^*$ is proportional to $k(\beta_{ij} - \beta_{ji})$; that is, if G is a game of strategic complements (substitutes), whoever exerts more concern towards his opponent will choose a larger (smaller) value of x. Next, since $2 + k(1 + \beta_{ij}) > 0$, (1) yields:[7]

$$\Pi_i \equiv \pi_i(x_i^*, x_j^*) = \frac{(2 + k(1 + \beta_{ij}))(2 + k(1 - \beta_{ij}(1 + k(1 + \beta_{ji}))))}{(4 - k^2(1 + \beta_{ij})(1 + \beta_{ji}))^2} > 0 \tag{2}$$

(Chapter 20 in Epple and Romano (2011))

[6] Following Levine (1998), the linearity of the subjective utility in the opponents material payoff is a convenient approximation. Bester and Güth (1998), Bolle (2000), Possajennikov (2000) and Carrasco et al. (2018) have used the same setting to study the evolutionary stability of interdependent preferences.

[7] To show that $\Pi_i > 0$ we do: as $4 - k^2(1+\beta_{ij})(1+\beta_{ji}) > 0$ then $\Pi_i(\beta_{ij}, \beta_{ji}) > 0 \leftrightarrow 2 + k - k\beta_{ij}(1 + k(1+\beta_{ji})) > 0$. If $0 < k \le 1$, as $-k(1+k(1+\beta_{ji})) < 0$ then $2+k-k\beta_{ij}(1+k(1+\beta_{ji})) > 2-(1+\beta_{ji})k^2 > 0$. If $-1 < k < 0$ then $-1 < 1 + k(1 + \beta_{ji}) \le 1$ and as $\beta_{ij} \in [-1, 1]$ then $-1 \le -\beta_{ij}(1 + k(1 + \beta_{ji})) \le 1$ so $-k(1 - \beta_{ij}(1 + k(1 + \beta_{ji}))) \le -2k < 2$.

Note that, exploiting (2), without accounting for intrinsic types nor peer effects, the unique equilibrium profile $\beta^*_{ij} = \beta^*_{ji} = k/(2-k)$ arises, as in Bester and Güth (1998) evolutionary approach. Crucially, at this point we replace their evolutionary process by a strategic interaction stage in which social influences (peers effects) impose a *match-specific restriction* that arise from $\beta_{ij} \equiv \theta_i + \lambda_i(\theta_j - \theta_i)$.[8] This solves the long run strategic interaction game at the preference level, restricting preferences in $\min(\theta_i, \theta_j) \leq \beta_{ij} \leq \max(\theta_i, \theta_j)$, allowing equilibria beyond the above Bester and Güth (1998) symmetric case.

Reciprocity (long run) Game: To explore peer effects we now compute the equilibrium reciprocity by solving game $\Lambda(k)$. Let $\kappa \equiv k/(2-k)$, and to avoid trivialities, we now assume that $\theta_i \neq \theta_j$ and $k \in (-1, 1)$, so $\kappa \in (-1/3, 1)$. When players choose how much reciprocity to exert, best responses are:[9]

$$\lambda_i(\lambda_j) = \frac{1}{(\theta_j - \theta_i)} \left(\frac{(1 + \beta_{ji}(\lambda_j))\kappa(1 + 2\kappa)}{(1+\kappa)^2 + \kappa(1 + \beta_{ji}(\lambda_j))} - \theta_i \right) \tag{3}$$

Standard derivation rules yields $\partial\lambda_i(\lambda_j)/\partial\lambda_j > 0$ if $\kappa < 0$ and $\partial\lambda_i(\lambda_j)/\partial\lambda_j < 0$ if $\kappa > 0$. That is, *when the short run game $G(\beta, k)$ is one of strategic complements (substitutes), the reciprocity game $\Lambda(k)$ becomes one of strategic substitutes (complements).* We omit the case $\kappa = -1/3$, as this was explored by Carrasco et al. (2018).

Proposition 1 *Peer effects exist in Λ.*

The proof of Proposition 1 argues that for any $(\theta_i, \theta_j) \in \Theta^2$ and any $\kappa \in (-1/3, 1)$, at least one player chooses a strictly positive reciprocity strategy in equilibrium. More specifically we show that if $(\theta_i - \theta_j)(\theta_j - \kappa) \geq 0$ then $\lambda^*_i = 1$ and $\lambda^*_j = 0$ is the unique Nash equilibrium; meaning that peer effects only arise in player i. Otherwise, reciprocity peer effects arise simultaneously in both players where $\lambda^*_i = (\theta_i - \kappa)/(\theta_i - \theta_j)$ and $\lambda^*_j = 1 - \lambda^*_i$ is the unique equilibrium. Observe that our result in Proposition 1 is independent of the value of κ, but the specific intensity choice of reciprocity might not. In fact, as κ rises, the more intrinsically altruistic player grow less reciprocal, whereas the opponent grows more reciprocal.

[8] In Bester and Güth (1998), an affine transformation of β_{ij} is inherited by each agent. An evolutionary process replaces our strategic interaction stage at the long run level. Unlike us, each player β parameter applies regardless of which agent it is matched with, meaning that concern in Bester and Güth (1998) is not match-specific. In this case player i best response is $\beta_{ij} = (1 + (2+k)k\beta_{ji})/(4 + (2-k)k(1+\beta_{ji}))$. Second order conditions for player i optimization is is $-k^2(4+(1+\beta_{ji})(2-k)k)^4/4(2+k(1+\bar{\beta}_{ji}))^2(2-k^2(1+\beta_{ji})) \leq 0$, and so equilibrium $\beta^*_{ij} = \beta^*_{ji} = k/(2-k)$.

[9] They arise by maximizing (2) in λ_i, subject to $\lambda_i \in [0, 1]$ and $\beta_{ij} \in [\min(\theta_i, \theta_j), \max(\theta_i, \theta_j)]$.

Note that altogether, in any equilibrium we have that $\lambda_i^* + \lambda_j^* = 1$, which guarantees that *peer effects exists at least in one player and that equilibrium preferences are symmetric*, so $\beta_{ij} = \beta_{ji}$.[10] In addition, despite the fact that players can be both selfish, induced preferences might not be as selfish as they could be. For instance, if $\kappa = 0.3$, $\theta_i = 0$ and $\theta_j = 0.2$, despite players could choose $\lambda_i = 0$ and $\lambda_j = 1$, in equilibrium they choose $\lambda_i^* = 1$ and $\lambda_j^* = 0$. Exploiting the inequalities statements of Proposition 1:

Corollary 1 *Induced preferences are:*

$$\beta_{ij}^*(\theta_i, \theta_j) = \beta_{ji}^*(\theta_j, \theta_i) = \min(\max(\kappa, \min(\theta_i, \theta_j)), \max(\theta_i, \theta_j))) \tag{4}$$

Regardless, the extreme cases of altruism or spitefulness where $\beta_{ij}^* = \beta_{ji}^* = 1$ or $\beta_{ij}^* = \beta_{ji}^* = -1$ are never induced interdependent preferences profiles.[11] In addition, selfish preferences $\beta_{ij}^* = \beta_{ji}^* = 0$ might arise, but only depending on the type of strategic interaction: if $\kappa < 0$ and $\min(\theta_i, \theta_j) = 0$ or $\kappa > 0$ and $\max(\theta_i, \theta_j) = 0$. Interestingly, in match, a single intrinsically selfish player is not sufficient to induce selfish preferences. Clearly, as preferences are a weighted average of players types, when both individuals are intrinsically altruists (spiteful) then they are also altruistic (spiteful). Otherwise, when players type significantly differ and only one is an altruist or only one is spiteful, the type of game that the individual play is relevant: altruistic (spiteful) preferences are only induced in games of strategic complements (substitutes). In this case, as depicted on the left panel of Figures 1 and 2 (gray regions), when $\min(\theta_i, \theta_j) < \kappa < \max(\theta_i, \theta_j)$, best responses yield a unique solution with $0 < \lambda_i^*, \lambda_j^* < 1$ and $\beta_{ij}^*(\theta_i, \theta_j) = \beta_{ji}^*(\theta_j, \theta_i) = \kappa$. Furthermore, *peer effects arise in both players* and the larger the complementarity (substitutability) of actions in the short run game, the more altruistic (spitefully) the induced preference is.

We now characterize the cases where *peer effects arises only in one player*. Graphically these cases are depicted on the left panel of Figures 1 and 2 (white regions). First, when both players types are sufficiently high and $\theta_i, \theta_j \geq \kappa$, only the highest type player chooses $\lambda^* = 1$ while the other is not reciprocal at all and $\lambda^* = 0$.[12] That is,

[10] In addition, reciprocity is not monotone neither in a players own type nor in the opponents type, by Proposition 1. In particular, it falls in θ_i when $\theta_i \leq \min(\kappa, \theta_j)$ and rises when $\theta_i \geq \max(\kappa, \theta_j)$; otherwise it equals zero, by Proposition 1. Equivalently, it is zero when $\theta_j < \min(\kappa, \theta_i)$ and jumps to one when $\min(\kappa, \theta_i) \leq \theta_j \leq \max(\kappa, \theta_i)$; otherwise falls in θ_j. As for the monotonicity in κ, we see that reciprocity λ_i is a piecewise linear function.

[11] As a result, concern for others yields inefficient outcomes. The efficient outcome arises when $\beta_{ij}^* = \beta_{ji}^* = 1$ and so $x_i^* = x_j^* = 1/4(1-k)$ for $k \neq 1$ and $x_i^* + x_j^* = 1/2$ for $k = 1$.

[12] When $\kappa > 0$ this requires players to be sufficiently intrinsically altruistic; when $\kappa < 0$ this requires players to be not too intrinsically spiteful.

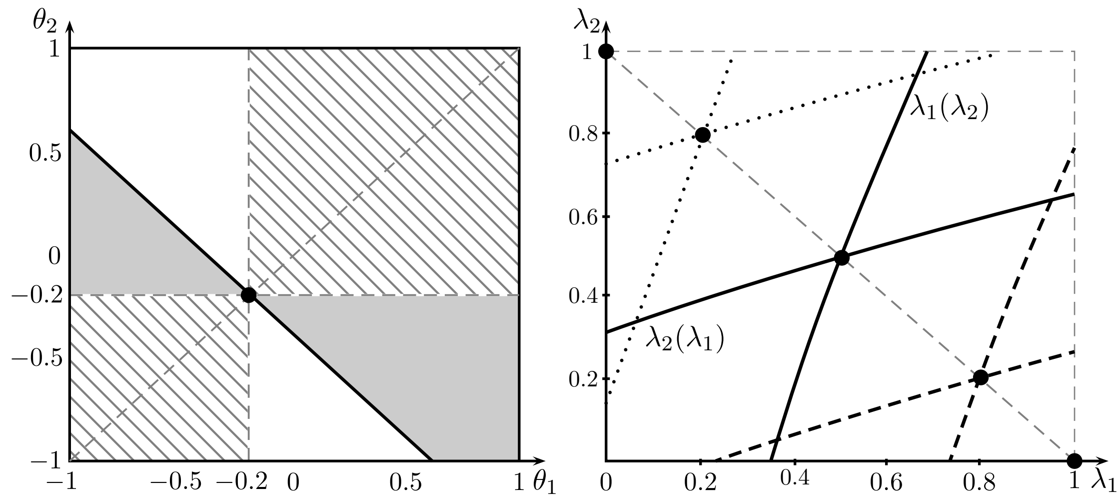

Figure 1: PREFERENCES AND RECIPROCITY FOR SUBSTITUTES ($\kappa = -1/5$). At left, in the upper (lower) dashed region the more (less) altruistic player exerts reciprocity; the other is not reciprocal. Equilibrium preferences are symmetric at $\beta^* = \min(\theta_1, \theta_2)$ in the upper white region and $\beta^* = \max(\theta_1, \theta_2)$ in the lower one. The gray L-shaped curves depict different preference values. In the remaining gray regions players behave spitefully at κ; the Bester and Güth (1998) equilibrium. At right, the reciprocity are complements and the best responses slope upwards. Any equilibrium obeys $\lambda_1^* + \lambda_2^* = 1$. We use (θ_1, θ_2) equal to $(0, -0.4)$ (solid), $(0, -0.25)$ (dashed) and $(-0.15, -0.4)$ (dotted).

peer effects arise only in the higher type player. In this case, the induced preferences are $\beta_{ij}^*(\theta_i, \theta_j) = \beta_{ji}^*(\theta_j, \theta_i) = \min(\theta_i, \theta_j)$, by (4). Equivalently, unlike spitefulness, altruism arises at their minimum possible intensity, at the lowest player type. Second, when players types are sufficiently low and $\theta_i, \theta_j \leq \kappa$, the lower type player chooses $\lambda^* = 1$ and the other $\lambda^* = 0$. That is, *peer effects arise only in the lower type player.* Induced preferences are $\beta_{ij}^*(\theta_i, \theta_j) = \beta_{ji}^*(\theta_j, \theta_i) = \max(\theta_i, \theta_j)$, by (4). To wit, unlike spite, altruism arises at the maximum possible intensity, at the highest players type. Altogether, preferences are restricted to be as close as κ.

Regardless, in equilibrium symmetric preferences are induced $\beta_{ij}^* = \beta_{ji}^* = \beta^*$ and so $\Pi_i = \Pi_j = (1+\kappa)(1+\kappa(1-2\beta^*))/4(1-\kappa\beta^*)^2$, by (2). That is, altruism increases each player payoff.[13] More specifically, when $-1/3 < \kappa < 1$, we have $x_i^* = x_j^* = 1/2(1-\kappa) > 0$ and $\Pi_i = \Pi_j = (1+2\kappa)/4(1-\kappa^2) > 0$. In this case, *equilibrium outcomes do not vary symmetrically in κ;*[14] while $\beta_{ij}^* = \beta_{ji}^* \to 1$, $x_i^* = x_j^* \to \infty$ and $\pi_i^* = \pi_j^* \to \infty$ as $\kappa \to 1$, we have that $\beta_{ij}^* = \beta_{ji}^* \to -1/3$, $x_i^* = x_j^* \to 3/8$ and $\pi_i^* = \pi_j^* \to 3/32$ as $\kappa \to -1/3$. On the other hand, we there is symmetric behavior, i.e $\lambda_i^* = \lambda_j^* = 1/2$ and $\beta_{ij}^* = \beta_{ji}^* = (\theta_i+\theta_j)/2$, only when $\theta_i + \theta_j = 2\kappa$, by Proposition 1. We depict the last two paragraphs analysis in Figure 1 — when G is a game of strategic substitutes — and Figure 2 — when G is

[13] As $\partial\Pi_i/\partial\beta^* = (1-\beta^*)(1+\kappa)\kappa^2/8(1-\kappa\beta^*)^3 > 0$, this extends Proposition 1 in Bester and Güth (1998). Altruism increases efficiency, whereas spite reduces it.

[14] This is a common characteristic in games of strategic complementaries and substitutes (Milgrom and Roberts (1990), Frankel et al. (2003))

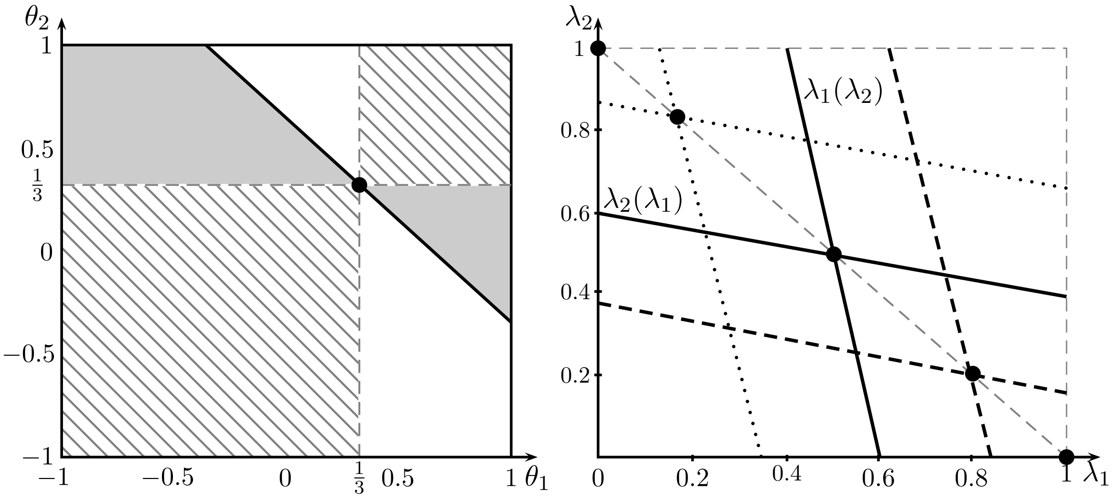

Figure 2: PREFERENCES AND RECIPROCITY FOR COMPLEMENTS ($\kappa = 1/3$). At left, in the upper (lower) dashed region the more (less) altruistic player exerts reciprocity; the other is not reciprocal. Equilibrium preferences are symmetric at $\beta^* = \min(\theta_1, \theta_2)$ in the upper white region and $\beta^* = \max(\theta_1, \theta_2)$ in the lower one. The gray L-shaped curves depict different preference values. In the remaining gray regions players behave spitefully at κ; the Bester and Güth (1998) equilibrium. At right, the reciprocity are substitutes and the best responses slope downwards. Any equilibrium obeys $\lambda_1^* + \lambda_2^* = 1$. We use (θ_1, θ_2) equal to $(1/2, 1/6)$ (solid),$(-0.3, 0.5)$ (dashed) and $(0.4, 0)$ (dotted).

a game of strategic complements.

So far we have characterized match-specific peer effects, that account for both players specific types. However it is equally interesting to understand what behavior is to be expected when players are under social influences (peer effect). We now further *quantify and sign peer effects* by exploring how much *expected preferences* $\bar{\beta}_i(\theta_i) = \int \beta_{ij}^*(\theta_i, \omega) dF(\omega)$ differ from their intrinsic value θ_i.[15] Rewrite (4):[16]

$$\bar{\beta}_i(\theta_i) = \begin{cases} \int \min(\theta_i, \max(\kappa, \omega)) dF(\omega) & \text{for } \theta_i \geq \kappa \\ \int \max(\theta_i, \min(\kappa, \omega)) dF(\omega) & \text{for } \theta_i \leq \kappa \end{cases} \tag{5}$$

Exploiting (5), and as depicted in Figure 3, we see that $\bar{\beta}_i(\theta_i)$ rises in θ_i and obeys $\bar{\beta}_i(\theta_i) \geq \theta_i$ iff $\theta_i \leq \kappa$ in which case *peer effects are positive*. It follows that a stochastically better opponents distribution leads to larger peer effects. In addition, $\bar{\beta}_i(\theta_i)$ is differentiable (except eventually at $\theta_i = \kappa$), is convex for $\theta_i \leq \kappa$ and concave for $\theta_i \geq \kappa$.[17] Furthermore, for extreme types, we have $\bar{\beta}_i(-1) = \mathbb{E}_{\theta_j}(\min(\kappa, \theta_j))$ and $\bar{\beta}_i(1) = \mathbb{E}_{\theta_j}(\max(\kappa, \theta_j))$, and so they vary differently as F grows more risky; while $\bar{\beta}_i(1)$ rises, $\bar{\beta}_i(-1)$ falls, by standard

[15]When $k = -1$, we restrict attention to the symmetric equilibrium, so that in all meetings preferences are symmetric and equal to $\beta_{ij}^*(\theta_i, \theta_j)$, as in (4).

[16]Write $\beta_{ij}^*(\theta_i, \theta_j) = \min(\max(\theta_i, \min(\kappa, \theta_j)), \max(\kappa, \theta_j)))$. As $\theta_i \geq \kappa \geq \min(\kappa, \theta_j)$, the first interval is obvious. Otherwise, since $\theta_i \leq \kappa$ and $\min(\kappa, \theta_j) < \max(\kappa, \theta_j)$, then $\beta_{ij}^*(\theta_i, \theta_j) = \max(\theta_i, \min(\kappa, \theta_j))$.

[17]We have $\partial\bar{\beta}_i(\theta_i)/\partial\theta_i = F(\kappa)$ if $\theta_i < \kappa$ and $\partial\bar{\beta}_i(\theta_i)/\partial\theta_i = 1 - F(\kappa)$ if $\theta_i > \kappa$. To with, the derivative at $\theta_i = \kappa$ only exists when $1 - F(\kappa) = F(\kappa)$.

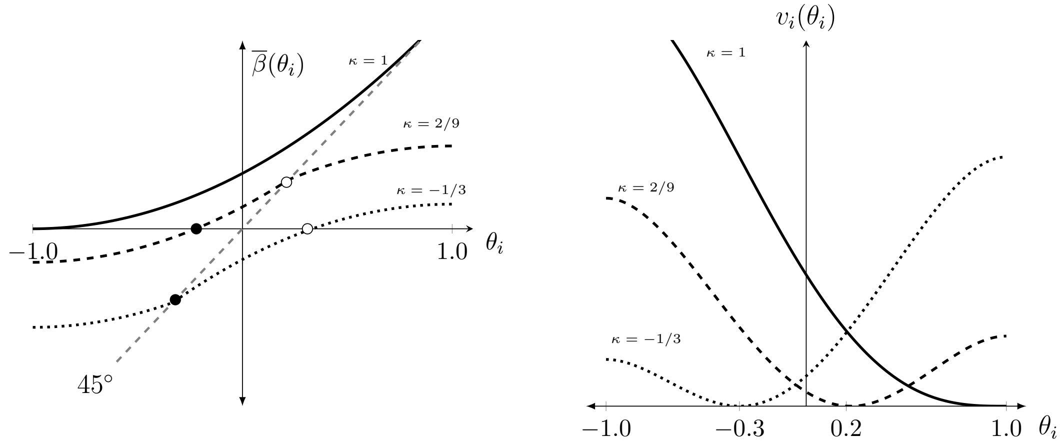

Figure 3: Interim Expected Preferences and its Variance. At left, the interim expected preferences. For $\theta_i \leq \underline{\theta}_i$ (black circles) then $\theta_i \leq \bar{\beta}_i(\theta_i) \leq 0$ and if $\theta_i \geq \bar{\theta}_i$ (white circles) then $0 \leq \bar{\beta}_i(\theta_i) \leq \theta_i$. Otherwise $\bar{\beta}_i(\theta_i) \leq \min(0, \theta_i)$ if $\kappa < 0$ and $\bar{\beta}_i(\theta_i) \geq \max(0, \theta_i)$ if $\kappa > 0$. At right, its variance. When $\theta_i = \kappa$, then $\beta^*_{ij} = \kappa$ and thus the variance is zero. We posit $\theta_i, \theta_j \sim u[-1, 1]$.

stochastic order logic. That is, extremely spiteful (altruistic) players grow more spiteful (altruistic) as the distribution is more dispersed.

Importantly, there might be cases where due to large peer effects, an intrinsically altruistic (spiteful) player expects to behave spitefully (altruistically). We call this phenomenon *preference-reversion*, whose existence is now established.

Proposition 2 *There is a nonempty set of types where preference-reversion arises in* Λ.

This result is based on the identification of critical values $-1 \leq \underline{\theta}_i < 0 < \bar{\theta}_i \leq 1$ such that if $\theta_i \in [\underline{\theta}_i, \bar{\theta}_i]$ then $\bar{\beta}_i(\theta_i) \geq \max(0, \theta_i)$ if $\kappa > 0$ and $\bar{\beta}_i(\theta_i) \leq \min(0, \theta_i)$ if $\kappa < 0$. That is, if the short run game is one of strategic substitutes (complements) then preference-reversion will arise only in moderately spiteful (altruistic) players.

We highlight three main conclusions derived from Proposition 2 and depicted on the left panel of Figure 3. First, expected preferences generically differ from the intrinsic types (only coincide in the particular case when $\theta_i = \kappa$), and so individuals generally restrict their own intrinsic preferences in equilibrium, translating into peer effects. Second, *peer effects intensity* reduces for extreme types: a sufficiently altruistic player $\theta_i \geq \bar{\theta}_i > 0$ expects to behave altruistically but not as much as he intrinsically is (i.e. peer effects are negative and $0 \leq \bar{\beta}_i(\theta_i) \leq \theta_i$), whereas a sufficiently spiteful player $\theta_i \leq \underline{\theta}_i < 0$ expects to behave spitefully, but not as much as he is (i.e. peer effects are positive and $\theta_i \leq \bar{\beta}_i(\theta_i) \leq 0$). Third, for those individuals that are neither sufficiently altruistic nor sufficiently spiteful we might observe that *preference-reversion* is induced— the fact that due to large peer effects, an intrinsically altruistic (spiteful) player expects to behave spitefully (altruistically); formally, that $\theta_i\bar{\beta}_i(\theta_i) < 0$. In these cases, when $\underline{\theta}_i < \theta_i < \bar{\theta}_i$, induced preferences might reverse depending on the strategic context: for $\kappa < 0$ ($\kappa > 0$) players expect to behave spitefully (altruistically).

We compute the *induced preferences variance* $v_i(\theta_i) = \mathbb{E}_{\theta_j}(\beta_{ij}^{*2}) - \bar{\beta}_i(\theta_i)^2$ using (4):

$$v_i(\theta_i) = \begin{cases} 2\int_\kappa^{\theta_i}(\theta_i - \omega)F(\omega)d\omega - \left(\int_\kappa^{\theta_i} F(\omega)d\omega\right)^2 & \text{for } \theta_i \geq \kappa \\ -2\int_{\theta_i}^{\kappa}(\omega - \kappa)F(\omega)d\omega - \left(\int_{\theta_i}^{\kappa} F(\omega)d\omega\right)^2 & \text{for } \theta_i \leq \kappa \end{cases} \quad (6)$$

As shown on the right panel of Figure 3. For $\theta_i < \kappa$, since $F(\omega) \leq 1$ and thus $\int_{\theta_i}^{\kappa} F(\omega)d\omega \leq \kappa - \theta_i$ we have $\partial v_i(\theta_i)/\partial\theta_i = 2F(\theta)\left(\theta_i - \kappa + \int_{\theta_i}^{\kappa} F(\omega)d\omega\right) < 0$. For $\kappa < \theta_i < 1$ we have $\partial v_i(\theta_i)/\partial\theta_i = 2(1 - F(\theta))\int_\kappa^{\theta_i} F(\omega)d\omega > 0$. At $\theta = \kappa$ then $v_i(\kappa) = 0$. That is, peer effects exhibit more variation when types differ more from κ. The variance depends not only on a player specific type, but also on the strategic of the short run game.

4 Peer Effects Under Incomplete Information

Depending on the assumption about the size of the group and type of encounter, perhaps it is more accurate to modify the information structure of the game. We now assume that the specific type of the opponent is unknown and that each player only knows the distribution. Our model also captures these strategic environment, as long as we relax the complete information assumption. Aiming for this, we now assume that in each meeting, each player i only knows the opponents group distribution F_j. As a result, they cannot condition their behavior on who they meet. Reciprocity strategies will only depend on each player own type. In any case, we use Section 3 logic to solve for equilibrium behavior.

Short run game: An immediate consequence of the incomplete information assumption is that now players cannot condition their behavior on who they meet. Instead, their induced preferences will only depend on each player own type. Let $b_i(\theta_i)$ be player i *incomplete information induced preference*, which is fixed at this stage. Players pairwise interaction defines a *normal form Bayesian game* $\mathcal{G} = \{\{i, j\}, (x_i, x_j) \in \mathbb{R}_+^2, \Theta^2, \{F, F\}, (u_i, u_j)\}$. The interim expected utility is:

$$U_i(x_i|\theta_i) = x_i(1 - x_i + k\mathbb{E}_{\theta_j}[x_j(\theta_j)]) + b_i(\theta_i)\mathbb{E}_{\theta_j}[x_j(\theta_j)(1 - x_j(\theta_j) + kx_i)] \quad (7)$$

Let $\bar{b}_j \equiv \mathbb{E}_{\theta_j}[b_j(\theta_j)]$. Then, solving for the Bayesian equilibrium strategies, we obtain:[18]

[18] Best responses are $x_i(\theta_i) = (1 + k(1 + b_i(\theta_i))\mathbb{E}_{\theta_j}[x_j(\theta_j)])/2$ and so $\mathbb{E}_{\theta_i}[x_i(\theta_i)] = (1 + k(1 + \bar{b}_i)\mathbb{E}_{\theta_j}[x_j(\theta_j)])/2$. By the same logic, we obtain $\mathbb{E}_{\theta_j}[x_j(\theta_j)]$and so $\mathbb{E}_{\theta_j}[x_j(\theta_j)] = (2 + k(1 + \bar{b}_j))/(4 -$

$$x_i^*(\theta_i) = \frac{(2 + k(b_i(\theta_i) - \bar{b}_i))(2 + k(1 + \bar{b}_j)) + 2k(\bar{b}_i - \bar{b}_j)}{2(4 - k^2(1 + \bar{b}_i)(1 + \bar{b}_j))} \tag{8}$$

The expected material payoffs $\mathbb{E}_{\theta_j}[\Pi_i] = x_i^*(\theta_i)(1 - x_i^*(\theta_i) + k\mathbb{E}_{\theta_j}[x_j^*(\theta_j)])$ are:

$$\mathbb{E}_{\theta_j}[\Pi_i] = \frac{(4 + k(2 - k\bar{b}_i(1 + \bar{b}_j)))^2 - k^2(2 + k(1 + \bar{b}_j))^2(b_i(\theta_i))^2}{4(4 - k^2(1 + \bar{b}_i)(1 + \bar{b}_j))^2} \tag{9}$$

Inspecting (9), we verify that absent any restriction on preferences we recover, as it is expected, Bester and Güth (1998) equilibrium and $b_i(\theta_i) = \bar{b}_i = k/(2 - k)$.

Reciprocity (long run) Game: To explore peer effects we now compute the equilibrium reciprocity by solving game $\Lambda(k)$—as in the complete information case— weighting their type and the average opponent's type. That is ($\star$) $b_i(\theta_i) = \theta_i + \lambda_i(\mu - \theta_i)$ and $\min(\theta_i, \mu) \leq b_i(\theta_i) \leq \max(\theta_i, \mu)$.

To solve for reciprocity strategy λ_i, we define the Bayesian game as $\Lambda = \{\{i, j\}, (\lambda_i, \lambda_j) \in [0, 1]^2, \Theta^2, \{F, F\}, (\mathbb{E}_{\theta_j}[\Pi_i], \mathbb{E}_{\theta_i}[\Pi_j])\}$ and maximize (9) in λ_i, accounting for ($\star$) and taking $\bar{b}_j$ as given. Exploiting our continuum type assumption we observe that the specific choice of λ_i does not modify the expected preference $\bar{b}_i$. Thus, inspecting (9) we can directly infer that expected material payoffs are maximized when induced preferences $b_i(\theta)$ are as close as zero as possible. That is, players reciprocity strategy aims for social preferences *as selfish as posible.* More formally,

Proposition 3 *Peer effects exists in Λ, unless $(\theta_i, \theta_j) \in [\min(0, \mu), \max(0, \mu)]^2$.*

Unlike the complete information case, now the existence of peer effects is not guaranteed. In fact, if $(\theta_i, \theta_j) \in [\min(0, \mu), \max(0, \mu)]^2$ then players are not reciprocal, peer effects does not arise, $\lambda_i^* = \lambda_j^* = 0$ and so preferences coincide with intrinsic values $b_i^*(\theta_i) = \theta_i$. Otherwise, peer effects arise, yet not necessarily in both players. More specifically, equilibrium reciprocity strategy for each player i is $\lambda_i^* = \theta_i/(\theta_i - \mu)$ if $\max(\theta_i, \mu) \geq 0 \geq \min(\theta_i, \mu)$, $\lambda_i^* = 1$ if $\theta_i > \mu \geq 0$ or $0 \geq \mu > \theta_i$ and $\lambda_i^* = 0$ if $0 \geq \theta_i > \mu$ or $\mu > \theta_i \geq 0$.

Note that if $\mu = 0$ then peer effects always exists.[19] Furthermore, $\lambda_i^* = \lambda_j^* = 1$ and so preferences are always selfish with $b_i^*(\theta_i) = b_j^*(\theta_j) = 0$. In any case, optimal reciprocity is not only independent of the opponents type, as one might expected, but also independent of the strategic context summarized in parameter k.

Using optimal reciprocity choices, we deduce that preferences are:

Corollary 2 *The interim expected preferences are:*

$k^2(1 + \bar{b}_i)(1 + \bar{b}_j))$. We obtain (8) by plugging $\mathbb{E}_{\theta_j}[x_j(\theta_j)]$ in player i best response.

[19] A sufficient condition is a symmetric F distribution.

$$b_i^*(\theta_i) = \min(\max(0, \min(\theta_i, \mu)), \max(\theta_i, \mu))) \quad (10)$$

For an intuition, observe that players recognize that their reciprocity decision cannot influence their opponents' behavior. Then, each player i maximizes $x_i(\theta_i)(A - x_i(\theta_i))$, where $A \equiv 1 + k\mathbb{E}_{\theta_j}[x_j(\theta_j)]$ acts as a *residual demand*, invariant to the reciprocity choice. It follows, by the well known monopoly rule, that $x_i(\theta_i) = A/2$ is optimal. Since player i equilibrium strategy in (8) dictates $x_i(\theta_i) = A/2 + k\mathbb{E}_{\theta_j}[x_j(\theta_j)]b_i(\theta_i)/2$, it is optimal to induce $b_i(\theta_i)$ as close to zero, and so preferences that are as selfish as possible.

Compared to (4), we see that preferences in the incomplete information case are as if players were to match the average opponents type μ when $\kappa \approx 0$. So, players restrict their preferences so that they behave as selfish as they can, given the constraints imposed by types. Easily, observe that $b_i^*(\theta_i) = 0$ whenever $\mu\theta_i \leq 0$; otherwise, if $\max(\theta_i, \mu) < 0$, spite is induced $b_i^*(\theta_i) < 0$ and if $\min(\theta_i, \mu) > 0$ they induce altruism $b_i^*(\theta_i) > 0$. That is, *a necessary and sufficient condition for altruism (spite) to arise is: both individuals intrinsic altruism (spitefulness) and altruistic (spiteful) expected opponents type*. Otherwise, selfishness arises as induced preferences. This result is at odds with the findings of Ely and Yilankaya (2001) who for a general model of indirect evolution show that with incomplete information — when the preferences of the opponent are not known — only egoistic preferences (or preferences equivalent to them) survive evolution.

Peer effects are now measured by $b_i^*(\theta_i) - \theta_i$. Exploiting the above inequalities we deduce that the conditions for positive or negative peer effects are different than those in our complete information model. In particular, *now there is no preference-reversion.*

Proposition 4 *There is no preference-reversion in* Λ.

The result shown by Proposition 4 is consistent with the fact that under a context of incomplete information, in which we assume that the specific type of the opponent is unknown and that each player only knows the distribution, the effects of social influence (peer effects) on individuals behaviors are somehow diluted. First, while Proposition 3 shows that in this setting peer effects do exist, they do so under certain conditions and indeed it is not guaranteed as it is in the complete information scenario. Consequently, also unlike the complete information scenario, Proposition 4 shows that preference-reversion is not possible, indicating that even under the presence of large peer effects, an intrinsically altruistic (spiteful) player will never change towards spitefully (altruistically)behavior.

5 Conclusions

We formalize the notion that people adjust their preferences and behavior influenced by the peers with whom they interact. Preferences are neither given, nor evolutionarily selected before they become givens. We present a model that accounts for pairwise random meetings between players with interdependent preferences who engage in a simultaneous move short-run game and engage in long-run strategic interaction at the preference level, where peer effects arise: they play an underlying reciprocity game, the equilibrium of which determines preferences and concern towards others. That is, reciprocity is our model's key ingredient for endogenous preferences and peer effects.

We measure peer effects by how much a player's induced preferences differ from his or her intrinsic values. We show that player reciprocity might differ leading to peer effects of different magnitudes. We provide evidence that reciprocity will induce preferences in some cases due to large peer effects. Intrinsically altruistic (spiteful) individuals expect to behave spitefully (altruistically); that is, that we might observe preference-reversion. Whether peer effects are positive or negative crucially depends on how types compare to the type of strategic interaction of the short-run game.

When there is incomplete information on other player types, assuming that the specific type of the opponent is unknown and that each player only knows the distribution, we show that there is no preference-reversion and that equilibrium preferences are as selfish as possible. Regardless, unlike Ely and Yilankaya (2001), altruistic and spiteful preferences might arise. Consequently, these results seem to suggest that the effects of social influence (peer effects) on individual behavior are somehow diluted under a context of incomplete information on other player types.

Future extensions to this work include replicating our analytical framework for more general matching technologies, other than pairwise random matching to examine the relationship between group size and peer effects. Finally, the theoretical results presented here can give rise to an experimental design to be tested empirically considering the impact of peer effects under scenarios of complete and incomplete information on other player types (for example, showing other players' previous experimental behaviors, see for instance Villena and Zecchetto (2011)).

6 Appendix

Proof of Proposition 1: Consider player $\theta_i \neq \theta_j$ maximization. As the Kuhn-Tucker FOC are necessary, we set up a Lagrangean $\mathcal{L} = \pi_i(x_i^*, x_j^*) + \gamma_0\lambda_i + \gamma_1(1-\lambda_i)$, where $\gamma_0, \gamma_1 \geq 0$ are the multipliers for $\lambda_i \geq 0$ and $\lambda_i \leq 1$. The FOC are:

$$\frac{(2+k(1+\beta_{ji}))((1+\beta_{ji})(2+k)k - \beta_{ij}(4+k(1+\beta_{ji})(2-k)))}{(4-k^2(1+\beta_{ij})(1+\beta_{ji}))^3} = \frac{(\gamma_1-\gamma_0)}{k^2(\theta_j-\theta_i)} \quad (11)$$

with $\gamma_0\lambda_i = 0$ and $\gamma_1(1-\lambda_i) = 0$. When $\lambda_i^* = 1$ and $\lambda_j^* = 0$ then $\gamma_0 = \gamma_1' = 0$, $\beta_{ij} = \beta_{ji} = \theta_j$ and $k^2(\theta_j - \theta_i)(k - \theta_j(2-k))/(2-k(1+\theta_j))^3(2+k(1+\theta_j)) = \gamma_1 = \gamma_0' \geq 0$, by (11). As $\theta_j \in [-1,1]$ then $(2-k(1+\theta_j))(2+k(1+\theta_j)) > 0$ and so $(\theta_j - \theta_i)(k - \theta_j(2-k)) \geq 0$.

For uniqueness, we argue that only $\lambda_i^* = 1$, $\lambda_j^* = 0$ solves the Kuhn-Tucker conditions, and as maximum exists in $[0,1]$, it is the unique maximum. We argue by contradiction, letting $(\theta_j - \theta_i)(k - \theta_j(2-k)) \geq 0$ and $\lambda_i^* < 1$ or $\lambda_j^* < 0$ or both. By Lemma 1, neither $\lambda_i = \lambda_j = 1$ nor $\lambda_i = \lambda_j = 0$ are equilibrium profiles.

Case 1: If $\lambda_i^* < 1$ and $\lambda_j^* > 0$: Then $\gamma_1 = \gamma_0' = 0$. If $\theta_i > \theta_j$ then $\theta_j \geq k/(2-k)$ and $\beta_{ij}, \beta_{ji} > k/(2-k)$. If $\theta_j > \theta_i$ then $\theta_j \leq k/(2-k)$ and $\beta_{ij}, \beta_{ji} < k/(2-k)$.

As $4 - k^2(1+\beta_{ij})(1+\beta_{ji}) > 0$ for $-1 < k < 1$, both players FOC in (11) yield:

$$(\theta_i - \theta_j)k(b_3 - \beta_{ij})(\beta_{ji} - b_1) \geq 0 \quad (12)$$

$$(\theta_i - \theta_j)k(b_4 - \beta_{ij})(\beta_{ji} - b_2) \geq 0 \quad (13)$$

with $b_1 = 4\beta_{ij}/(k(2+k-\beta_{ij}(2-k))) - 1$, $b_2 = (1+\beta_{ij})(2+k)k/(4+k(1+\beta_{ij})(2-k))$, $b_3 = (2+k)/(2-k)$, $b_4 = -4/k(2-k) - 1$ and:

$$\frac{(b_2-b_1)(b_3-\beta_{ij})(b_4-\beta_{ij})}{(\beta_{ij}(2-k)-k)} = \frac{4(1+k)(2+k(1+\beta_{ij}))}{k^2(2-k)^2} \geq 0 \quad (14)$$

Observe that $b_3 > k/(2-k)$, $b_4 > b_3 \leftrightarrow k < 0$ and $b_4 > k/(2-k) \leftrightarrow k < 0$.

For $k < 0$ and $\theta_i > \theta_j$, then $b_4 > b_3 > k/(2-k)$ and $\beta_{ij}, \beta_{ji} > k/(2-k)$. If $b_3 > \beta_{ij} > k/(2-k)$ then (12), (13) and (14) yield $\beta_{ji} \leq b_1$. But as $\beta_{ij} = b_1$ at $\beta_{ij} = k/(2-k)$ and $\partial b_1/\partial\beta_{ij} = 4(2+k)/k(2+k-\beta_{ij}(2-k))^2 < 0$, then for $\beta_{ij} > k/(2-k)$ we have $\beta_{ji} < k/(2-k)$. A contradiction. If $b_4 > \beta_{ij} > b_3$ then (12), (13) and (14) yield $b_1 \leq \beta_{ji} \leq b_2$ and $b_2 \leq b_1$. A contradiction. If $\beta_{ij} > b_4$ then (12), (13) and (14) yield $\beta_{ji} \geq b_2$. But since $\beta_{ij} = b_2$ at $\beta_{ij} = k/(2-k)$ and $\partial b_2/\partial\beta_{ij} = 4(2+k)k/(4+k(1+\beta_{ij})(2-k))^2 < 0$, then for $\beta_{ij} > k/(2-k)$ we have $\beta_{ji} < k/(2-k)$. A contradiction.

If $k < 0$ and $\theta_j > \theta_i$, then $\beta_{ij}, \beta_{ji} < k/(2-k) < b_3 < b_4$. To wit (12), (13) and

(14) dictate $\beta_{ji} \geq b_1$. But as $\beta_{ij} = b_1$ at $\beta_{ij} = k/(2-k)$ and $b_1 = \partial b_1/\partial \beta_{ij} < 0$, then $\beta_{ij} < k/(2-k)$ yield $\beta_{ji} > k/(2-k)$. A contradiction.

If $0 < k < 1$ then $b_3 > 1$ and $b_4 < -1$ so $b_2 \geq b_1$ iff $\beta_{ij} \leq k/(2-k)$. Now (12) and (13) dictate $(\theta_j - \theta_i)(\beta_{ji} - b_1) \leq 0$ and $(\theta_j - \theta_i)(\beta_{ji} - b_2) \geq 0$. When $\theta_i > \theta_j$ this reduces to $b_1 \leq \beta_{ji} \leq b_2$, and $b_1 \geq b_2$, by (14). A contradiction. Equivalently, if $\theta_j > \theta_i$ this reduces to $b_2 \leq \beta_{ji} \leq b_1$ and $b_2 \geq b_1$. A contradiction.

CASE 2: IF $\lambda_i^* < 1$ AND $\lambda_j^* = 0$: Then $\gamma_1 = \gamma_1' = 0$, $\beta_{ji} = \theta_j$, $\theta_j < \beta_{ij} \leq \theta_i$ if $\theta_i > \theta_j$ and $\theta_i \leq \beta_{ij} < \theta_j$ if $\theta_j > \theta_i$. In this case the FOC yield (12) and the reversed inequality of (13). We now use the same logic of the previous case. For $k < 0$ and $\theta_i > \theta_j$, if $b_3 > \beta_{ij} > k/(2-k)$ then (12), (13) and (14) yield $b_2 \leq \beta_{ji} \leq b_1$ and $b_1 \leq b_2$. A contradiction. If $b_4 > \beta_{ij} > b_3$ then (12), (13) and (14) yield $\beta_{ji} \geq b_1$. But $\partial b_1/\partial \beta_{ij} < 0$ and $b_1 > 1$ at $\beta_{ij} = b_4$, so for $b_4 > \beta_{ij} > b_3$ we have $\beta_{ji} > 1$. A contradiction. If $\beta_{ij} > b_4$ then (12), (13) and (14) yield $b_1 \leq \beta_{ji} \leq b_2$ and $b_2 \geq b_1$. But since $\beta_{ij} = b_2$ at $\beta_{ij} = k/(2-k)$ and $\partial b_2/\partial \beta_{ij} < 0$, then $\beta_{ij} > k/(2-k)$ yields $\beta_{ji} < k/(2-k)$. A contradiction. For $k < 0$ and $\theta_j > \theta_i$, then $\beta_{ij}, \beta_{ji} < k/(2-k) < b_3 < b_4$. To wit (12), (13) and (14) dictate $b_1 \leq \beta_{ji} \leq b_2$ and $b_2 \leq b_1$. A contradiction.

If $0 < k < 1$ then $(\theta_j - \theta_i)(\beta_{ji} - b_1) \leq 0$ and $(\theta_j - \theta_i)(\beta_{ji} - b_2) \leq 0$ by (12) and (13) and $b_2 \leq b_1 \leftrightarrow \beta_{ij} \geq k/(2-k)$ by (14). When $\theta_i > \theta_j$ then $\beta_{ij} > k/(2-k)$, so the FOC reduce to $\beta_{ji} \geq b_1$. But $b_1 - \beta_{ij} = -(2+k(1+\beta_{ij}))(k - \beta_{ij}(2-k))/k(2+k-\beta_{ij}(2-k))$, then $\beta_{ij} \geq b_1 \leftrightarrow \beta_{ij} \leq k/(2-k)$. To wit, $\beta_{ij} < b_1 \leq \beta_{ji}$. A contradiction. Equivalently, if $\theta_j > \theta_i$ then $\beta_{ij} < k/(2-k)$ so the FOC yield $\beta_{ji} \leq b_1$. To wit, $\beta_{ji} \leq b_1 < \beta_{ij}$. A contradiction.

CASE 3: IF $\lambda_i^* = 1$ AND $\lambda_j^* > 0$: Then $\gamma_0 = \gamma_0' = 0$, $\beta_{ij} = \theta_j$, $\theta_j < \beta_{ji} \leq \theta_i$ if $\theta_i > \theta_j$ and $\theta_i \leq \beta_{ji} < \theta_j$ if $\theta_j > \theta_i$. In this case the FOC yield the reversed inequality of (12) and (13). We now use the same logic of the previous case. For $k < 0$ and $\theta_i > \theta_j$, if $b_3 > \beta_{ij} > k/(2-k)$ then (12), (13) and (14) yield $b_1 \leq \beta_{ji} \leq b_2$ and $b_1 \leq b_2$. But since $b_2 = \beta_{ij}$ at $\beta_{ij} = k/(2-k)$ and $\partial b_2/\partial \beta_{ij} < 0$, then $\beta_{ij} > k/(2-k)$ yields $\beta_{ji} < k/(2-k)$. A contradiction. If $b_4 > \beta_{ij} > b_3$ then (12), (13) and (14) yield $b_2 \leq \beta_{ji} \leq b_1$ and $b_2 \leq b_1$. But as $b_1 = \beta_{ij}$ at $\beta_{ij} = k/(2-k)$ and $\partial b_1/\partial \beta_{ij} < 0$, then for $\beta_{ij} > k/(2-k)$ we have $\beta_{ji} < k/(2-k)$. A contradiction. If $\beta_{ij} > b_4$ then (12), (13) and (14) yield $b_2 \leq \beta_{ij} \leq b_1$ and $b_1 \geq b_2$. A contradiction.

If $k < 0$ and $\theta_j > \theta_i$, then $\beta_{ij}, \beta_{ji} < k/(2-k) < b_3 < b_4$. To wit (12), (13) and (14) dictate $b_2 \leq \beta_{ji} \leq b_1$ and $b_2 \leq b_1$. But as $b_1 = \beta_{ij}$ at $\beta_{ij} = k/(2-k)$ and $\partial b_2/\partial \beta_{ij} < 0$, then for $\beta_{ij} < k/(2-k)$ we have $\beta_{ji} > k/(2-k)$. A contradiction.

If $0 < k < 1$ then (12) and (13) dictate $(\theta_j - \theta_i)(\beta_{ji} - b_1) \geq 0$ and $(\theta_j - \theta_i)(\beta_{ji} - b_2) \geq 0$. When $\theta_i > \theta_j$ this reduces to $\beta_{ji} \leq b_2$, as $\beta_{ij} > k/(2-k)$. But $b_2 - \beta_{ij} = (2 + k(1 +$

$\beta_{ij}))(k-\beta_{ij}(2-k))/(4+k(1+\beta_{ij})(2-k))$, then $\beta_{ij} \leq b_2 \leftrightarrow \beta_{ij} \leq k/(2-k)$. To wit, $\beta_{ji} \leq b_2 < \beta_{ij}$. A contradiction. If $\theta_j > \theta_i$ then $\beta_{ji} \geq b_2$. To wit, $\beta_{ij} < b_2 \leq \beta_{ji}$. A contradiction.

For the interior equilibrium, we intersect best responses in (3). This yields two candidates for equilibrium: $\lambda = (\theta_i - \kappa)/(\theta_i - \theta_j)$ and $\lambda' = (2+\theta_i+1/\kappa)/(\theta_i-\theta_j)$. We discard λ' as it is either negative or exceeds one. Letting $\lambda \in (0,1)$ yields $(\theta_i - \theta_j)(\theta_j - \kappa) < 0$. Behavior is $\beta^*_{ij} = \beta^*_{ji} = \kappa$.□

Proof of Proposition 2: Integrating (5) by parts yields:

$$\bar{\beta}_i(\theta_i) = \begin{cases} \theta_i - \int_{\kappa}^{\theta_i} F(\omega)d\omega & \text{for } \theta_i \geq \kappa \\ \kappa - \int_{\theta_i}^{\kappa} F(\omega)d\omega & \text{for } \theta_i \leq \kappa \end{cases} \tag{15}$$

We now find the critical values $\underline{\theta}_i$ and $\bar{\theta}_i$ such that if $\theta_i \in [\underline{\theta}_i, \bar{\theta}_i]$ then preference-reversion arises. We divide the analysis into cases.

For $\kappa > 0$: as $\partial\bar{\beta}_i(\theta_i)/\partial\theta_i = 1 - F(\theta_i) \in [0,1)$ if $\theta_i \geq \kappa$, by (15), then $0 \leq \bar{\beta}_i(\theta_i) \leq \theta_i$. Hence, $\bar{\theta} = \kappa$. If $\theta_i \leq \kappa$ then $\bar{\beta}_i(\theta_i) \geq \theta_i$ and $\partial\bar{\beta}_i(\theta_i)/\partial\theta_i = F(\theta_i) \in [0,1)$. If $\bar{\beta}_i(-1) = \mathbb{E}(\min(\kappa, \Theta_j)) < 0$ then a unique $\underline{\theta}_i > -1$ solves $\bar{\beta}_i(\underline{\theta}_i) = 0$ and so $\theta_i \leq \bar{\beta}_i(\theta_i) \leq 0$ for $\theta_i \leq \underline{\theta}_i$. Easily, as $\bar{\beta}_i(0) > 0$, then $\underline{\theta}_i < 0$. If $\bar{\beta}_i(-1) \geq 0$ then $\underline{\theta}_i = -1$. Easily, $\bar{\beta}_i(\theta_i) \geq \max(0, \theta_i)$ if $\underline{\theta}_i \leq \theta_i \leq \bar{\theta}_i$.

For $\kappa < 0$: if $\theta_i \leq \kappa$ then $\theta_i \leq \bar{\beta}_i(\theta_i) \leq 0$ and so $\underline{\theta}_i = \kappa$, by (15). Next, for $\theta_i \geq \kappa$ we have $\partial\bar{\beta}_i(\theta_i)/\partial\theta_i = 1 - F(\theta_i) \in [0,1)$ and $\bar{\beta}_i(0) < 0$, by (15). To wit, a unique $\bar{\theta}_i > 0$ solves $\bar{\beta}_i(\bar{\theta}_i) = 0$ and for $\theta_i \geq \bar{\theta}_i$ then $0 \leq \bar{\beta}_i(\theta_i) \leq \theta_i$. We have $\bar{\theta}_i < 1$ iff $\bar{\beta}_i(1) = \mathbb{E}(\max(\kappa, \Theta_j)) > 0$. Easily, $\bar{\beta}_i(\theta_i) \leq \min(0, \theta_i)$ if $\underline{\theta}_i \leq \theta_i \leq \bar{\theta}_i$.□

Proof of Proposition 3: We optimize (9) in λ_i. Observe that given our linear specification for preferences, $\partial b_i(\theta_i)/\partial\lambda_i = \mu_j - \theta_i$. Fix $\bar{b}_j$ and $\bar{b}_i$; then, computing the FOC we get:

$$\frac{\partial \mathbb{E}_{\theta_j}[\Pi_i]}{\partial \lambda_i} = \frac{-k^2(2+k(1+\bar{b}_j))^2 b_i^*(\theta_i)(\bar{\theta}-\theta_i)}{2(4-k^2(1+\bar{b}_i)(1+\bar{b}_j))^2} = 0 \leftrightarrow b_i^*(\theta_i) = 0$$

Clearly, if $\theta_i = \mu_j$ then any λ_i is optimal. Otherwise, the solution is $\lambda_i^* = \theta_i/(\theta_i - \mu_j)$. To guarantee $0 \leq \lambda_i^* \leq 1$ we restrict types to $\max(\theta_i, \mu_j) \geq 0 \geq \min(\theta_i, \mu_j)$. Otherwise, if $\theta_i > \mu_j \geq 0$ or $0 \geq \bar{\theta} > \theta_i$ then $\lambda_i^* = 1$, and $\lambda_i^* = 0$ if $0 \geq \theta_i > \mu_j$ or $\mu_j > \theta_i \geq 0$. Altogether, $\lambda_i^* > 0$ or $\lambda_j^* > 0$ always, unless $(\theta_i, \theta_j) \in [\min(0,\mu), \max(0,\mu)]^2$. □

Proof of Proposition 4: Assume $\theta_i\mu \neq 0$. Then, we see that if $\theta_i < \min(0,\mu)$ peer effects are positive and negative if $\theta_i > \max(0,\mu)$. That is, a sufficiently spiteful player expect to behave spitefully, but not as much as he is (i.e. $0 > b_i^*(\theta_i) > \theta_i$), whereas a sufficiently altruistic player expects to behave altruistically but not as much as he intrinsically is (i.e. $\theta_i > b_i^*(\theta_i) > 0$). Otherwise — when $\min(0, \mu) < \theta_i < \max(0, \mu)$ — peer effects might be positive of negative depending on the sign of μ. Regardless, there is no preference-reversion. Whenever $\mu\theta_i > 0$, then $\theta_i > b_i^*(\theta_i) > 0$ and so peer effects are negative. When $\mu\theta_i < 0$, then $0 > b_i^*(\theta_i) > \theta_i$ and so peer effects are positive. □

Lemma 1 *For $k \neq 0$, neither $\lambda_i = \lambda_j = 1$ nor $\lambda_i = \lambda_j = 0$ are equilibrium profiles.*

Proof: We show that no $\theta_i, \theta_j \in [-1, 1]$ solve simultaneously the FOC in (11), which are necessary for a maximum. In either case, the FOC dictate:

$$(\theta_i - \theta_j)(4\theta_j - k(1 + \theta_i)(2 - k)(b_3 - \theta_j)) \geq 0 \tag{16}$$

$$(\theta_j - \theta_i)(4\theta_i - k(1 + \theta_j)(2 - k)(b_3 - \theta_i)) \geq 0 \tag{17}$$

If $\theta_i > \theta_j$ then $k(1 + \theta_i)(2 - k)(b_3 - \theta_j) \leq 4\theta_j < 4\theta_i \leq k(1 + \theta_j)(2 - k)(b_3 - \theta_i)$ and so $k(2 - k)(\theta_i - \theta_j)(b_3 + 1) < 0$; a contradiction for $k > 0$. The $\theta_j > \theta_i$ case is analogous. If $k < 0$ and $\theta_i > \theta_j$ then $\theta_i \in [0, b_3]$ clearly does not solve (17). Neither does $\theta_i \in (b_3, 1]$, as $k(1 + \theta_j)(2 - k)(b_3 - \theta_i)$ rises linearly from 0 to $2k^2(1 + \theta_j) < 4k^2 < 4$. The only candidates are $\theta_j < \theta_i < 0$. But in this case the FOC yield:

$$\frac{k(1 + \theta_i)(2 + k)}{4 + k(1 + \theta_i)(2 - k)} \leq \theta_j \leq \frac{4\theta_i}{k(2 - k)(b_3 - \theta_i)} - 1 \tag{18}$$

This inequality limits are equal at $\theta_i = k/(2 - k)$. Easily, their slopes are negative, and the upper limit slope exceeds the lower limit slope iff $-2 - 2\theta_i k + k^2 + \theta_i k^2 \geq 0$, which is iff $-k(2 - k)(\theta_i + (2 - k^2)/k(2 - k)) \geq 0$. A contradiction. To wit, the interval in (18) is empty. The $\theta_j > \theta_i$ case is analogous. □

References

Alger, I. and J. W. Weibull (2013): "Homo Moralis–Preference Evolution Under Incomplete Information and Assortative Matching," *Econometrica*, 81, 2269–2302.

Arnott, R. and J. Rowse (1987): "Peer group effects and educational attainment," *Journal of Public Economics*, 32, 287–305.

Bandiera, O., I. Barankay, and I. Rasul (2010): "Social Incentives in the Workplace," *The Review of Economic Studies*, 77, 417–458.

Bardsley, N. and R. Sausgruber (2005): "Conformity and reciprocity in public good provision," *Journal of Economic Psychology*, 26, 664 – 681.

Benabou, R. (1993): "Workings of a city: location, education, and production," *The Quarterly Journal of Economics*, 108, 619–652.

Beshears, J., J. J. Choi, D. Laibson, B. C. Madrian, and K. L. Milkman (2015): "The effect of providing peer information on retirement savings decisions," *The Journal of finance*, 70, 1161–1201.

Bester, H. and W. Güth (1998): "Is altruism evolutionarily stable?" *Journal of Economic Behavior and Organization*, 34, 193 – 209.

Bolle, F. (2000): "Is altruism evolutionarily stable? And envy and malevolence?: Remarks on Bester and Güth," *Journal of Economic Behavior and Organization*, 42, 131–133.

Bougheas, S., J. Nieboer, and M. Sefton (2013): "Risk-taking in social settings: Group and peer effects," *Journal of Economic Behavior & Organization*, 92, 273 – 283.

Carrasco, J. A., R. Harrison, and M. Villena (2018): "Interdependent preferences and endogenous reciprocity," *Journal of Behavioral and Experimental Economics*, 76, 68 – 75.

Charness, G. and M. Rabin (2002): "Understanding Social Preferences with Simple Tests," *The Quarterly Journal of Economics*, 117, 817–869.

Chen, Y., F. M. Harper, J. Konstan, and S. X. Li (2010): "Social comparisons and contributions to online communities: A field experiment on movielens," *American Economic Review*, 100, 1358–98.

Dong, B., U. Dulleck, and B. Torgler (2012): "Conditional corruption," *Journal of Economic Psychology*, 33, 609 – 627.

Duflo, E. and E. Saez (2002): "Participation and investment decisions in a retirement plan: the influence of colleagues' choices," *Journal of Public Economics*, 85, 121 – 148.

Dufwenberg, M. and G. Kirchsteiger (2019): "Modelling kindness," *Journal of Economic Behavior & Organization*, 167, 228–234.

Ely, J. C. and O. Yilankaya (2001): "Nash equilibrium and the evolution of preferences," *journal of Economic Theory*, 97, 255–272.

Epple, D. and R. E. Romano (2011): "Peer effects in education: A survey of the theory and evidence," in *Handbook of social economics*, Elsevier, vol. 1, 1053–1163.

Falk, A., U. Fischbacher, and S. Gächter (2013): "Living in two neighborhoods?social interaction effects in the laboratory," *Economic Inquiry*, 51, 563–578.

Falk, A. and A. Ichino (2006a): "Clean evidence on peer effects," *Journal of labor economics*, 24, 39–57.

——— (2006b): "Clean Evidence on Peer Effects," *Journal of Labor Economics*, 24, 39–57.

Fehr, E. and K. M. Schmidt (1999): "A Theory of Fairness, Competition, and Cooperation," *The Quarterly Journal of Economics*, 114, 817–868.

Fershtman, C. and U. Segal (2018): "Preferences and Social Influence," *American Economic Journal: Microeconomics*, 10, 124–42.

Frankel, D. M., S. Morris, and A. Pauzner (2003): "Equilibrium selection in global games with strategic complementarities," *Journal of Economic Theory*, 108, 1 – 44.

Frey, B. S. and S. Meier (2004): "Social comparisons and pro-social behavior: Testing" conditional cooperation" in a field experiment," *American Economic Review*, 94, 1717–1722.

Gächter, S., D. Nosenzo, and M. Sefton (2012): "The impact of social comparisons on reciprocity," *The Scandinavian Journal of Economics*, 114, 1346–1367.

——— (2013): "Peer effects in pro-social behavior: Social norms or social preferences?" *Journal of the European Economic Association*, 11, 548–573.

Gächter, S. and C. Thöni (2010): "Social comparison and performance: Experimental evidence on the fair wage–effort hypothesis," *Journal of Economic Behavior & Organization*, 76, 531–543.

Gaviria, A. and S. Raphael (2001): "School-Based Peer Effects and Juvenile Behavior," *The Review of Economics and Statistics*, 83, 257–268.

Güth, W. and S. Napel (2006): "Inequality Aversion in a Variety of Games – An Indirect Evolutionary Analysis*," *The Economic Journal*, 116, 1037–1056.

Guryan, J., K. Kroft, and M. J. Notowidigdo (2009): "Peer effects in the workplace: Evidence from random groupings in professional golf tournaments," *American Economic Journal: Applied Economics*, 1, 34–68.

Ho, T.-H. and X. Su (2009): "Peer-induced fairness in games," *American Economic Review*, 99, 2022–49.

Ichino, A. and G. Maggi (2000): "Work environment and individual background: Explaining regional shirking differentials in a large Italian firm," *The Quarterly Journal of Economics*, 115, 1057–1090.

Isoni, A. and R. Sugden (2019): "Reciprocity and the Paradox of Trust in psychological game theory," *Journal of Economic Behavior & Organization*, 167, 219–227.

Koçkesen, L., E. A. Ok, and R. Sethi (2000): "The Strategic Advantage of Negatively Interdependent Preferences," *Journal of Economic Theory*, 92, 274 – 299.

Kremer, M. and D. Levy (2008): "Peer effects and alcohol use among college students," *Journal of Economic perspectives*, 22, 189–206.

Lazear, E. P. (2001): "Educational production," *The Quarterly Journal of Economics*, 116, 777–803.

Levine, D. K. (1998): "Modeling Altruism and Spitefulness in Experiments," *Review of Economic Dynamics*, 1, 593 – 622.

Mas, A. and E. Moretti (2009): "Peers at work," *American Economic Review*, 99, 112–45.

McMillan, R. (2004): "Competition, incentives, and public school productivity," *Journal of Public Economics*, 88, 1871–1892.

Milgrom, P. and J. Roberts (1990): "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities," *Econometrica*, 58, 1255–1277.

Mittone, L. and M. Ploner (2011): "Peer pressure, social spillovers, and reciprocity: an experimental analysis," *Experimental Economics*, 14, 203–222.

Okuno-Fujiwara, M. and A. Postlewaite (1995): "Social norms and random matching games," *Games and Economic behavior*, 9, 79–109.

Possajennikov, A. (2000): "On the evolutionary stability of altruistic and spiteful preferences," *Journal of Economic Behavior and Organization*, 42, 125 – 129.

Sacerdote, B. (2001): "Peer effects with random assignment: Results for Dartmouth roommates," *The Quarterly journal of economics*, 116, 681–704.

Sobel, J. (2005): "Interdependent preferences and reciprocity," *Journal of economic literature*, 43, 392–436.

Thöni, C. and S. Gächter (2015): "Peer effects and social preferences in voluntary cooperation: A theoretical and experimental analysis," *Journal of Economic Psychology*, 48, 72–88.

Villena, M. G. and F. Zecchetto (2011): "Subject-specific performance information can worsen the tragedy of the commons: Experimental evidence," *Journal of Economic Psychology*, 32, 330–347.

Waldinger, F. (2011): "Peer effects in science: Evidence from the dismissal of scientists in Nazi Germany," *The Review of Economic Studies*, 79, 838–861.

Zenou, Y., E. Patacchini, and E. Rainone (2014): "Heterogeneous Peer Effects in Education," .

Zimmerman, D. J. (2003): "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *The Review of Economics and Statistics*, 85, 9–23.