Strategic ethics: Altruism without the other-regarding confound

Giuseppe Attanasi * Kene Boun My † Nikolaos Georgantzís ‡ Miguel Ginés §

GREDEG Working Paper No. 2019–13

Abstract

In a two-stage investment-effort game, we model altruistic investment in another agent's capacity to benefit from synergies between the two agents' efforts. Contrary to most models in the literature on altruism, we assume that agents who invest in others have no direct utility from their giving behavior, ruling out any genuinely altruistic component in their utility function, *i.e.*, stemming from other-regarding preferences. Furthermore, we disentangle this "strategic ethics" from reputational effects yielding incentives for a more pro-social action in the present in order to favor Pareto-superior outcomes in the future.

Isolated consumption of one's own benefits from own efforts is the worst equilibrium, which is globally stable and is shown to exist independently of the investment cost. However, for a low enough investment cost, there exist two alternative equilibria: an unstable intermediate equilibrium in which both agents make positive complementarity-building investments, and a stable one in which both agents invest all they can to complementarity building. Both equilibria Pareto-dominate the aforementioned no-investment equilibrium.

Results of a laboratory experiment confirm our behavioral prediction that, for a low enough investment cost, subjects coordinate on positive complementarity-building investment, which in turn boosts their effort in the second stage. The latter increases in both own and others' complementarity-building investment, as predicted by our model. All this holds independently of subjects' risk and inequity aversion. The latter suggests that complementarity-building investment is not motivated by altruism. Rather, it is purely strategic.

Keywords: Complementarity-building Investment, Strategic Complementarities, Altruism, Fairness, Risk Aversion.

JEL classification: C72; C73; C91; D64.

*Corresponding author. GREDEG, Université Côte d'Azur (giuseppe.attanasi@unice.fr). We thank for useful discussions and comments Pierpaolo Battigalli, Luca Corazzini, Elena Manzoni, Salvatore Vergine, and the seminar participants at University of Florence in 2018, the 2013 Economic Science Association (ESA) World Meeting, and the 2013 Società Italiana degli Economisti (SIE) Annual Conference. The research leading to these results has received funding from the French Agence Nationale de la Recherche (ANR), under grant ANR-18-CE26-0018 (project GRICRIS).

†BETA, University of Strasbourg

‡Burgundy School of Business & Economics Department, Universitat Jaume I

§Economics Department, Universitat Jaume I

1 Introduction

Altruism is a human attitude involving a sacrifice in own wealth or utility in order to increase wealth or utility of another person. Most economic theories of altruism explain the sacrifice of own wealth in favor of another person by reducing altruistic behavior into (selfish) utility-increasing actions. However, some actions like the transmission of knowledge among co-authors of a scientific paper or of culture by mentors to advisees might correspond to purely altruistic, not own utility-increasing investment, which strategically help another person to benefit from both agents' actions.

It is increasingly remembered nowadays that Adam Smith in his Theory of Moral Sentiments (1759) had explicitly referred to the pleasure of contemplating others' happiness. This fact alone can explain altruistic behavior. The formalization of this type of altruism has led to approaches which, by complying with the axioms of neoclassical economics, tend to reduce altruism down to the standard story of selfish (own) utility maximization. This is due to the fact that the existing neoclassical approaches to altruism need some (positive) other-regarding components in an individual's utility function to explain why an individual's own actions may benefit the society surrounding him/her.

Even if we take a person's genuine altruistic motivation as a fact, it is not easy to explain why we observe investments aimed at increasing others' efforts in the future, like cultural transmission from parents to their children, from older to young generations, from the society to "newcomers" like immigrants and students from abroad. Of course, some conservative attitudes aimed at irrationally preserving the local culture from external inputs and drastic changes could help us explain cultural transmission as an irrational adherence to what "has been there since centuries ago". But modern and progressive cultural messages are also transmitted even in the least advanced societies, in which case a rational approach would be necessary.

In this paper, we propose a framework in which each player can be both a transmitter and a receiver. A strategic investment is made in the first stage by each "transmitter" to his/her "receiver" in order for the latter to improve his/her capacity to benefit from synergies emerging due to both agents' efforts made in the second stage. In other words, we envisage such an investment as a way of improving each "receiver's" ability to understand, communicate, improve, appreciate and finally directly benefit from the interaction with his/her "sender", rendering their mutual efforts synergic, in order for a higher effort to be made by each one in a "strategically altruistic" equilibrium. We also identify an alternative equilibrium in which no investment takes place. In that case, individuals enjoy the benefits from their efforts in isolation, choosing lower effort levels and yielding lower levels of social welfare.

Certainly closer to the selfish approach to altruistic behavior, our framework does not include any other-regarding component in an agent's utility function. In this, it differs from both distribution-dependent models of altruism (Fehr and Schmidt 1999, Bolton and Ockenfels 2000, Charness and Rabin 2002) and intention-based models of reciprocity (Rabin 1993, Charness and Rabin 2002, Dufwenberg and Kirchsteiger 2004, Falk and Fischbacher 2006, Cox et al. 2008).[1] Nor does our model depend on the repetition

[1]For a review of several of these models, see Attanasi and Boun My (2016).

of human interaction. Rather, it focuses on a one-shot game. In fact, in the control treatment of the experimental setup, we rule out the possibility of strategic behavior à la Andreoni (1988), by focusing on randomly rematched individuals in each period of the game (stranger matching). Therefore, there is no reputational effect yielding incentives for a more pro-social action in the present in order to favor Pareto superior outcomes in the future. In another treatment, we test the interplay between strategic complementarity building and strategic reputation building, by focusing on a setting in which individuals' initial matching is maintained in each period of the game (partner matching).

Independently of the matching, in each period of our investment-effort game the resulting behavior looks more like reciprocity of the type "I invest on you, in order to increase your effort in a task which is synergic to my actions", with the novel feature that the magnitude of strategic complementarity is endogenously determined by agents' actions.

Our work contributes to several strands of the literature which aim at justifying actions that benefit others.

First of all, Suetens (2005) and Potters and Suetens (2009) have studied experimental settings in which strategic complementarity and substitutability mediate in the synergic effects from the two players' efforts. However, in these works interaction between efforts is not endogenously determined by altruistic investments, as is the case in our framework.

Second, reciprocally altruistic equilibria have been shown to emerge as dynamically stable strategies in a finitely repeated game (e.g., conditional cooperation in a repeated prisoner, public-good or other social dilemma), especially by rewarding or punishing co-players with monetary or non-monetary mechanisms (see Masclet et al. 2003, Masclet and Villeval 2008, and Masclet et al. 2013). However, as Villeval (2012) states, exploring the role of the heterogeneity of social preferences in the decision to cooperate in repeated social dilemmas reveal that conditional cooperation explains the decay of cooperation over time. As the comparison between our stranger and partner treatments shows, "conditional cooperation" in our complementarity-building game does not require repetition: if it is "strategically ethic" (*i.e.*, the investment cost is low enough), it occurs even in the last periods of both treatments, resembling one of the equilibria of the stage game.

Finally, the literature on biological altruism has proved that the evolution of a species selects the most altruistic genes. Three strands of this literature are potentially relevant to the goal of our paper: evolution of cooperation through population dynamics (Axelrod 1984, Ale et al. 2013); genetically-predisposed reciprocal altruism (Trivers 1971); and, cultural transmission (Bisin and Verdier 2000).

Especially the latter strand of studies has offered a rationale of why altruistic actions are easier to observe when agents act sequentially. In fact, the experimental literature on cultural transmission is mainly characterized by laboratory studies with different "generations" of subjects, where each group of subjects represents a generation, and is replaced by another group (generation) when the former group finishes playing and leaves the laboratory. Among these studies, the most relevant to the goal of our paper are those by Schotter and Sopher (2003, 2006, 2007) on two-player games (respectively, Battle of the Sexes, Trust, and Ultimatum Game) where, within the same experimental session, the first generation transmits to the next generation its history of play and/or advice about how to play the game. Independently of the role in the game, the first generation of two players always has an explicit monetary interest towards the next generation.[2] In all of these games, Schotter and Sopher find that word-of-mouth social learning (in the form of advice from laboratory "parents" to laboratory "children") can be a strong force in the creation of social conventions. In particular, when the game played by each generation is the Ultimatum Game (Schotter and Sopher 2007), subjects appear to follow conventions of reciprocity in that they tend to send more if they think the receivers acted in a "kind" manner, where kind means the senders sent more money than the receiver expected.[3]

Differently from all these experiments, in our model "cultural transmission" is modeled within a unique generation as an altruistic complementarity-building investment toward the other agent. The game between the "parents" and the "children" is then one-shot and symmetric: each player is both parent and son of the other. Among numerous examples of strategic transmission of culture that our study may explain there is co-authorship in a paper where each co-author has a different expertise, *e.g.*, a game theorist and an econometrician, with each author investing his/her time at the beginning of the interaction to "teach" the other the technicalities of his/her own contribution to the paper. This should ultimately increase the effort each co-author will put in the joint research.

Theoretical analysis of our game shows that isolated consumption of one's own benefits from own efforts is the worst equilibrium ("bad" equilibrium), which unfortunately for the society is globally stable and is shown to exist in all cases. On the opposite extreme there may be an alternative equilibrium in which all agents invest all they can in complementary-building. This equilibrium ("good" equilibrium), when it exists, is also stable and Pareto dominates the aforementioned "bad", no-investment equilibrium. Out-of-equilibrium behavior involves potentially asymmetric and "unfair" configurations in which the agent who invests more in cultural transmission ends up with lower utility.[4]

Results of an unframed laboratory experiment confirm our behavioral prediction that for a low enough investment cost subjects coordinate on positive complementarity-building investments, which in turn boost their effort levels. The latter increase in both own and other complementarity-building investment, as predicted by our model. All this holds independently of subjects' social preferences. This suggests that complementarity-building investment is not motivated by altruism. Rather, it is purely strategic.

The paper is structured as follows. Section 2 presents the model, and Section 3 describes the experimental design. Section 4 reports and discusses the empirical findings, and Section 5 concludes.

[2]For intergenerational games with no explicit monetary interest of the first generation or with an interest in conflict with that of the second one, see, respectively, Chaudhuri et al. (2006) and Kuang et al. (2007). More recently, Attanasi et al. (2017) have investigated experimentally role models with transmission of behavior and identity (and no advice and monetary interest) from a first generation of graduate students to a second generation of undergraduate students, one year after the font session (with graduate students).

[3]However, Charness and Villeval (2009) have shown in both laboratory and field experiments that agents' willingness to cooperate and to compete is mainly affected by the generation of the group members, even without seniors playing before juniors and/or "transmitting" them anything.

[4]A model in which in equilibrium cooperative individuals have a lower utility and invest more in cultural transmission has been provided by Della Lena et al. (2019), who show that high-guilt types in an environment with incomplete information of the matching may have a lower utility but still invest more in the transmission of their cooperative type to future generations in the presence of imperfect empathy.

2 A model of "altruistic" complementarity-building investments

2.1 Assumptions and notation

Two players, $i \in \{1, 2\}$ decide on the level of their **effort** x_i, which yields them a linear benefit and a quadratic cost. Apart from these effects of effort on own utility, the two players' efforts may interact positively to yield a further benefit to each one of them. However, each player's capacity to benefit from the interaction of efforts depends on the other's fully altruistic decision to undertake a costly **investment** in the former's synergy-absorbing capacity. This investment is altruistic in the sense that it is costly and has no direct positive effect on the *altruist*'s utility. This is captured by a utility function like the one given by:

$$U_i = x_i + \frac{\beta_{-i}}{1+\beta_{-i}} x_i x_{-i} - \frac{1}{2} x_i^2 - C(\beta_i), \tag{1}$$

where β_i is firm i's investment in $-i$'s (the opponent firm's) capacity to benefit from synergies arising due to the two agents' efforts x_i,$x_{-i} \in \Re_+$ and $C(\cdot) \geq 0$ is the investment cost function (the same for both firms), with $C'(\cdot) \geq 0$ and $C''(\cdot) \geq 0$.

It should be noted that the player's investment in the other's synergy-absorbing capacity has solely a negative direct impact on own utility, and a potentially (if efforts are both higher than 0) positive impact on the other's utility. The specification of effort synergies in (1) is similar to the way in which spill-overs are usually modeled in IO models of R&D competition. There, several results show that firms may wish to share with other their technological advances[5], but none of the models has gone as far as to assume that a firm may want to undertake costly actions to increase the other's capacity to absorb the synergies arising from their R&D efforts. Despite the obvious analogies, the literature on spill-overs between firms has not been sufficiently exploited to provide explanations for reciprocal behavior between individuals or social groups. Strangely, Vives' (1990, 2005, 2009) work on strategic complementarities has also remained unexploited by theorists of strategic interaction between humans. In fact, Vives (2009) explicitly refers to the way in which multistage interaction may affect equilibria in the presence of strategic complementarities. In this sense, our model may be seen as a special case of Vives' (2006) framework, although, from a technical point of view some of our assumptions contradict those in Vives (2005). Also, we are particularly interested in equilibria which are different from the obvious candidate of non-investment in others' complementarity-absorbing capacity.

2.2 Equilibria of the game

The two players play a two-stage game once. In the first stage, players decide simultaneously on their *gift*-investment in each other's synergy-absorbing capacity. In the second

[5] For example, Gil-Moltó et al. (2005) show that firms may deliberately choose to increase their technological similarities in order to increase their mutual benefits from synergies arising from their R&D efforts. Furthermore, Milliou (2006), in an endogenous spill-over duopoly model, shows that firms would make no costly investment to protect their innovations, which is a minimum requirement for the main result presented here to hold.

stage, they simultaneously decide on their effort levels.

Following backward induction, we first discuss the unique Nash equilibrium of the **second stage** of the game (effort levels).

Setting the derivative of (1) with respect to the effort level x_i equal to zero, we obtain the first order conditions:

$$\frac{\partial U_i}{\partial x_i} = 0 \Rightarrow x_i = 1 + \frac{\beta_{-i}}{1+\beta_{-i}} x_{-i}. \tag{2}$$

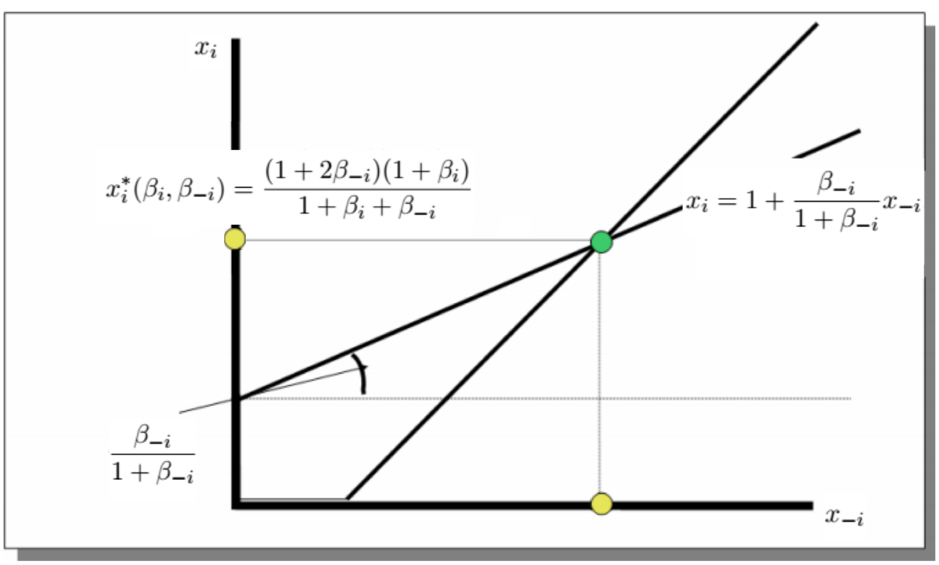
Note that the slope of i's effort-reaction function in (2) positively depends (approaching unity asymptotically from below) on $-i$'s altruistic investment in the former's capacity to absorb the synergies from the interaction of efforts. Thus, player i's reaction to $-i$'s effort will positively depend on $-i$'s investment in the former's synergy-absorbing capacity. Our theoretical results (Results 2 and 3 below) depend on this property of interaction in the effort-choice stage of the game. It is also worth noting that efforts are strategic complements only for strictly positive values of β. In the absence of effort synergies, strategic interaction in effort levels disappears, yielding effort levels of 1 and equilibrium utilities of 1/2.

Solving the system of reaction functions in eq. (2) with respect to x_i, x_{-i} gives equilibrium effort levels as a function of the two players' investment levels in the first stage of the game:

$$x_i^*(\beta_i, \beta_{-i}) = \frac{(1+2\beta_{-i})(1+\beta_i)}{1+\beta_i+\beta_{-i}}. \tag{3}$$

Effort-reaction functions in eq. (2) and equilibrium effort levels in eq. (3) are represented in Figure 1.

Figure 1: Best response functions and equilibrium in the effort stage.



Substituting eq. (3) into eq. (1), after some nontrivial but standard re-arrangements, we find:

$$U_i^*(\beta_i,\beta_{-i}) = \frac{(1+2\beta_{-i})^2(1+\beta_i)^2}{2(1+\beta_i+\beta_{-i})^2} - C(\beta_i). \tag{4}$$

We use this expression to solve the **first stage** of the game (investment levels). Differentiating (4) with respect to β_i gives:

$$\frac{\partial U_i^*(\beta_i,\beta_{-i})}{\partial \beta_i} = \frac{\beta_{-i}(1+2\beta_{-i})^2(1+\beta_i)}{(1+\beta_i+\beta_{-i})^3} - \frac{\partial C(\beta_i)}{\partial \beta_i}. \tag{5}$$

Further differentiating the first term of the right hand side expression of (5), we obtain that the equilibrium utility net of altruistic costs is neither convex nor concave as implied by:

$$\frac{\partial\left(\frac{\beta_{-i}(1+2\beta_{-i})^2(1+\beta_i)}{(1+\beta_i+\beta_{-i})^2}\right)}{\partial \beta_i} = \frac{\beta_{-i}(1+2\beta_{-i})^2(\beta_{-i}-2-2\beta_i)}{(1+\beta_i+\beta_{-i})^4} \geq 0. \tag{6}$$

Thus, having assumed that $C(\cdot)$ is convex, we can see that, generally speaking, the equation:

$$\frac{\beta_{-i}(1+2\beta_{-i})^2(1+\beta_i)}{(1+\beta_i+\beta_{-i})^3} = \frac{\partial C}{\partial \beta_i} \tag{7}$$

may not lead to an interior equilibrium in the first stage of the game. We illustrate here the existence of such equilibria and explore their properties adopting the following specification of convex cost function of the investment:

$$C(\beta_i) = \frac{1}{2}k_i((1+\beta_i)^2-1), \tag{8}$$

where $k_i > 0$ is a parameter measuring the convexity of the investment cost function. In fact, $\frac{\partial^2 C}{\partial \beta_i^2} = k_i$.

Then, substituting eq. (8) into the right-hand side of eq. (7), the latter can be rewritten as:

$$\frac{\beta_{-i}(1+2\beta_{-i})^2(1+\beta_i)}{(1+\beta_i+\beta_{-i})^3} = k_i(1+\beta_i), \tag{9}$$

or, simplifying:

$$\frac{\beta_{-i}(1+2\beta_{-i})^2}{(1+\beta_i+\beta_{-i})^3} = k_i. \tag{10}$$

Now, from eq. (10) it is easy to derive a unique best-response function in altruistic investments on the positive reals, using Descartes' sign rule of a polynomial:

$$\beta_i(\beta_{-i}) = \max\{0, \sqrt[3]{\frac{(1+2\beta_{-i})^2\beta_{-i}}{k_i}} - \beta_{-i} - 1\}. \tag{11}$$

This corresponds to a maximum, since $\beta < \beta_i(\beta_{-i})$ implies $\frac{\partial U_i^*(\beta,\beta_{-i})}{\partial \beta_i} > 0$ and $\beta > \beta_i(\beta_{-i})$ implie $\frac{\partial U_i^*(\beta,\beta_{-i})}{\partial \beta_i} < 0$.

Next, we discuss five properties of the best-response functions described by eq. (11), which are graphically represented together with the resulting equilibria in the two panels of Figure 2.

Property 1: β_i is an increasing function of β_{-i} if the concavity of the investment cost function is not extremely high, *i.e.*, $k_i \leq 4$ (this can be easily shown by deriving eq. (11) with respect to β_{-i}).

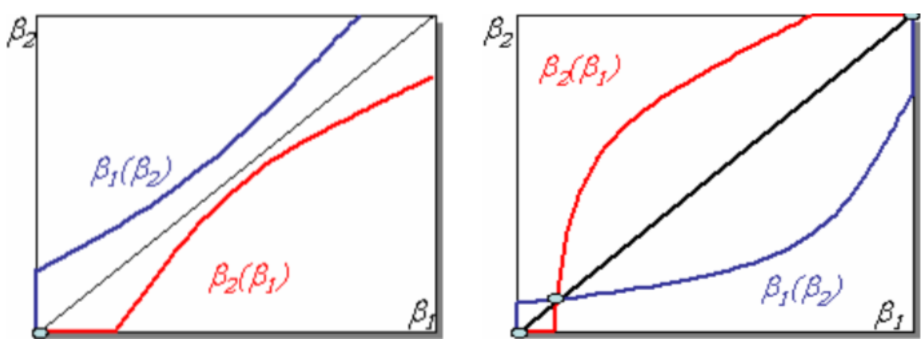
Property 2: $\beta_i(\beta - i)$ is a concave function independently from k_i, the concavity of the investment cost function (this can be shown after some computation on the derivatives of eq. (11)).

Property 3: If $k_i \leq 4$ (*i.e.*, β_i increasing in β_{-i}), then there exists a lower bound on the co-player's investment $\underline{\beta}_{-i}$ such that $\beta_i(\underline{\beta}_{-i}) = 0$ and $\beta > \underline{\beta}_{-i}$ implies $\beta_i(\beta) > 0$, *i.e.*, the player's best-response function is positive.

Property 4: For a high enough concavity of the investment cost function (*i.e.*, $k_i \geq 1/2$), $\beta_i(\beta)$ does not cross the diagonal, *i.e.*, there does not exist an equilibrium with positive complementary-building investment. Otherwise, $k_i < 1/2$ implies that $\widehat{\beta} = \frac{k_i}{1-2k_i}$ is the point where the function $\beta_i(\beta)$ crosses the diagonal (interior equilibrium).

Property 5: For a low enough concavity of the investment cost function (*i.e.*, $k_i < 1/2$), if there exists an upper bound $\overline{\beta} > \widehat{\beta} = \frac{k_i}{1-2k_i}$ on the possible β, then there also exists an upper bound on the co-player's investment $\overline{\beta}_{-i} \leq \overline{\beta}$ such that $\beta_i(\overline{\beta}_{-i}) = \overline{\beta}$.

Figure 2: Best response functions and equilibrium in the investment stage.



$k_1 = k_2 > 0.5$ $\qquad$ $k_1 = k_2 < 0.5$

With all this in mind, the first result we can derive is that the *bad* equilibrium always exists independently from k_i, the concavity of the investment cost function.

Result 1: The Bad equilibrium $(\beta_1^*, \beta_2^*) = (0, 0)$ exists for all $k_1, k_2 > 0$.

Proof: If firm $-i$ sets $\beta_{-i} = 0$, the right-hand side of eq. (7) is always larger than the left hand side ($k_i > 0$). Thus, firm i's best response will also be $\beta_i = 0$. By symmetry, this implies that this is always an equilibrium.□

Note that, if $k_1, k_2 \geq 1/2$, then no equilibrium other than the *bad* one described above exists, since none of the players' best response crosses the diagonal (see Property 4 above). This case is depicted in the left panel of Figure 2. Conversely, for $k_1, k_2 < 1/2$, there are other two equilibria besides the *bad* one: an *interior* one and the *good* one, which implies full reciprocity.

Result 2 concerns the existence of an interior equilibrium of the game.

Result 2: An Interior equilibrium (β_1^*, β_2^*) exists for all $k_1, k_2 < 1/2$.

Proof: Given $k_i < 1/2$ for all $i = 1, 2$, we know that $\beta(\overline{\beta}) > \overline{\beta}$ and the interior equilibrium is the solution to the following system of equations (from eq. (11)):

$$\begin{cases} \beta_1(\beta_2) = \max\{0, \sqrt[3]{\frac{(1+2\beta_2)^2\beta_2}{k_1}} - \beta_2 - 1\} \\ \beta_2(\beta_1) = \max\{0, \sqrt[3]{\frac{(1+2\beta_1)^2\beta_1}{k_2}} - \beta_1 - 1\} \end{cases}$$

Define the following function: $S(\beta) = \beta - \sqrt[3]{\frac{(1+2\beta')^2\beta'}{k_1}} + \beta' + 1$ with $\beta' = \sqrt[3]{\frac{(1+2\beta)^2\beta}{k_2}} - \beta - 1$. $S(0) = \sqrt[3]{\frac{1}{k_1}}$ and $S(\overline{\beta}) < 0$ since β' is increasing in β and $\beta'(\overline{\beta}) > \overline{\beta}$. Then, by continuity, there exists a β^* such that $S(\beta^*) = 0$ and $\beta_1^* = \beta^*$ and $\beta_2^* = \beta_2(\beta_1^*)$. $\square$

Resut 3 concerns the existence of a good equilibrium of the game.

Result 3: If there is a bound on the maximal investment $\overline{\beta}$ for each agent, a Good equilibrium (β_1^*, β_2^*) exists for all $k_1, k_2 < 1/2$.

Proof: If there is a bound on the maximal investment $\overline{\beta}$ for each agent, by Property 5 the good equilibrium is characterized by agents investing their full capacity.$\square$

Thus, as anticipated above, for a low enough convexity of the investment cost, *i.e.*, $k_i < 1/2$ for all $i = 1, 2$, we find three equilibria, the *bad* one (Result 1), an *interior* one (Result 2), and the *good* one, which implies full reciprocity (Result 3). The three aforementioned equilibria are represented in the right panel of Figure 2 and can be easily ranked using the Pareto criterion, with the bad equilibrium being Pareto dominated by both the internal and the good equilibrium. A straightforward proof is available from the authors of the fact that: when unique, the equilibrium is also globally stable; when good and bad equilibria coexist, they are both locally stable; interior equilibria are unstable.

2.3 Behavioral hypotheses

To elaborate behavioral hypotheses for an experimental test, we need to parametrize the players' utility function in eq. (1) and the investment cost function in eq. (8). To achieve saliency, in the experimental game we rescale players' payoff function U_i in eq. (1) by 10 times and express it in euros. Therefore, player $i \in \{1, 2\}$ in each experimental pair maximizes:

$$U_i = 10\left(x_i + \frac{\beta_{-i}}{1+\beta_{-i}}x_i x_{-i} - \frac{1}{2}x_i^2 - \frac{1}{2}k_i((1+\beta_i)^2 - 1)\right), \qquad (12)$$

with $\beta_i, \beta_{-i} \in [0, \overline{\beta}]$ and $x_i, x_{-i} \in [0, \overline{x}]$, $\overline{\beta} = \overline{x} = 3$, and 0.10 grids for both complementary-building investments and efforts.[6]

Note that the cost function in eq. (12) is the same as the one in eq. (8). The main treatment variable is the value we assign to $k_i = k_{-i}$, the concavity of the investment cost function in eqs. (8) and (12). We have two main treatments: **High-cost**, with $\mathbf{k_i = k_{-i} = 0.6}$, and **Low-cost**, with $\mathbf{k_i = k_{-i} = 0.4}$. Noting that the game with rescaled payoff function in eq. (12) is strategically equivalent to the game with the utility function described in eq. (1), we know that: there is a bad equilibrium in both treatments $(\beta^*_{bad}, x^*_{bad})$ (Result 1); in the Low-cost treatment there are also an interior equilibrium $(\beta^*_{int}, x^*_{int})$ (Result 2) and a good equilibrium $(\beta^*_{good}, x^*_{good})$ (Result 3). The equilibrium predictions with risk-neutral and selfish agents in the two treatments are reported in Table 1. The internal equilibrium values for the complementary-building investment have been calculated according to eq. (11), and the equilibrium values for the effort in the three equilibria have been calculated by substituting respectively β^*_{bad}, β^*_{int}, and β^*_{good} into eq. (3).

Table 1: Equilibrium predictions with risk-neutral and selfish agents

Treatment	**Bad** Equil.			**Interior** Equil.			**Good** Equil.		
k	β^*_{bad}	x^*_{bad}	U^*_{bad}	β^*_{int}	x^*_{int}	U^*_{int}	β^*_{good}	x^*_{good}	U^*_{good}
0.6	0.0	1.0	5.0						
0.4	0.0	1.0	5.0	2.0	3.0	29.0	3.0	3.0	22.5

Acknowledging that a standard equilibrium analysis like the one we provide in Section 2.2 has no compelling foundation for games played one-shot and in experiments on other-regarding preferences,[7] in each experimental session we repeat the two-stage game of Section 2.1 parametrized as in eq. (12) for 25 periods. In fact, in a repeated game the equilibria shown in Table 1 can be justified as stable states of learning dynamics.[8] In particular, recall that the bad and the good equilibria of Table 1 are stable, while the interior one is not.

For each of the two main treatments shown in Table 1, we implement two variations according to the matching protocol, ending up with four treatments (see Section 3.2 for specific implementation features). In the two **Stranger** treatments, the two-stage game of Section 2.1 is repeated 25 times under a stranger matching: new pairs are formed at

[6]We choose the set of investments and the set of efforts to be of the same size to avoid potential distortions in players' choice (*e.g.*, choosing $\beta_i < x_i$ due to an upper bound on the maximal investment $\overline{\beta}$ lower than the upper bound on maximal effort $\overline{x}$).

[7]See, *e.g.*, Attanasi et al. (2019b), using best-reply analysis rather than equilibrium analysis in a two-player embezzlement game with selfish, inequity-averse, or guilt-averse players.

[8]See, *e.g.*, Attanasi et al. (2019a), implementing a 12-period repeated game with stranger matching in order to test their equilibrium predictions with selfish and inequity-averse players

the beginning of each new period. In the two **Partner** treatments, it is repeated 25 times under a partner matching: pairs are kept the same in each of the 25 periods.

We elaborate our first behavioral hypothesis by relying on eq. (3), which describes how a player's equilibrium effort level in the second stage depends on the observed levels of investment of the pair in the first stage. We introduce this hypothesis as the first one since it is the crucial element of our model: players' investments in the first stage are complementarity-building, in the sense that they induce higher efforts in the second stage.

Behavioral Hypothesis 1 [Dependence of effort on investments] A player's effort level positively depends on both his own investment level and on the co-player's investment level. This holds independently of the treatment.

Next, we analyze treatment effects. The equilibrium predictions of Table 1 hold in each of the 25 periods of the two Stranger treatments: $(\beta^*_{bad}, x^*_{bad})$ in the High-cost treatment; $(\beta^*_{bad}, x^*_{bad})$, $(\beta^*_{int}, x^*_{int})$, or $(\beta^*_{good}, x^*_{good})$ in the Low-cost treatment. Therefore, we expect a higher cooperation among players – in terms of both complementarity-building investments and effort – in the Low-cost than in the High-cost treatment.

In the two Partner treatments, we know from the standard theory of repeated games (see Battigalli 2019) that: in the High-cost treatment the unique subgame-perfect Nash Equilibrium is given by the repetition of $(\beta^*_{bad}, x^*_{bad})$ across the 25 periods; in the Low-cost treatment there is a multiplicity of subgame-perfect Nash Equilibria, given by the play of any of the three equilibria in each of the 25 periods of the game. Therefore, our model predicts the same positive difference in investment and effort in favor of the Low-cost treatment. However, strategic pro-social behavior in the first periods and its decline across time is a common phenomenon in repeated social dilemma games played with partner matching (see, *e.g.*, Andreoni 1988 and a plethora of follow-up papers). Therefore, we expect strategic reputation building in both Partner treatments, especially in the High-Cost treatment. In fact, while in the High-Cost treatment strategic reputation building would just lead to play the good equilibrium more often than predicted, in the Low-Cost treatment it would lead to play $(\beta_{int}, x_{int}) = (2, 3)$ or $(\beta_{good}, x_{good}) = (3, 3)$ rather than $(\beta^*_{bad}, x^*_{bad}) = (0, 1)$, which would lead players a payoff of, respectively, $U_{int} = 21$ and $U_{good} = 7.5$ rather than the predicted $U^*_{bad} = 5$. Despite that, we presume that the main strategic feature of our game – complementarity-building investment – will survive the confound of "strategic" reputation building phenomenon, which will only have a second-order effect. Therefore, although milder than in the Stranger treatments, we expect a positive difference between the Low-cost and the High-cost treatment in terms of both complementarity-building investment and effort.

All this leads us to elaborate two behavioral hypotheses on treatment effects.

Behavioral Hypothesis 2 [High- vs. Low-cost]: In the Stranger treatments, the levels of investment and effort are significantly higher in the Low-cost than in the High-cost treatment. This difference is less significant in the Partner treatments and especially in the first periods of the repeated game.

Behavioral Hypothesis 3 [Stranger vs. partner matching]: In the High-cost treatments, the levels of investment and effort are significantly higher in the Partner than in the Stranger treatment. This difference is less significant in the Low-cost treatments and especially in the first periods of the repeated game.

Let us now relax the assumption of risk neutrality and selfish preferences.

It is easy to show that the qualitative features of the equilibrium predictions in Table 1 still hold if players are risk-averse, a characteristic that we elicit in our experiment (see Game 2 in Section 3.2). In fact, if both players are risk-averse, only the bad equilibrium emerges for $k > 0.5$ (High-cost treatments), and the bad, an interior, and the good equilibrium emerge for $k < 0.5$ (Low-cost treatments). Furthermore, the interior and the good equilibrium would still Pareto dominate the bad one, although the positive spread between the former and the latter would shrink, being lower, the higher the degree of players' risk aversion in the pair. Therefore, Behavioral Hypotheses 2 and 3 above can be extended to players non-neutral to risk if the following behavioral hypothesis is verified.[9]

Behavioral Hypothesis 4 [Independence from risk aversion]: The levels of investment and effort are independent from the player's degree of risk aversion. This holds in each treatment.

As for other-regarding preferences, we must distinguish between different forms of distributive and belief-dependent preferences, since different types of social preferences may have different strategic impact on players' behavior in the one-shot and the repeated social dilemmas (Villeval 2012). As already stated above, in this study we try to challenge distribution-dependent models of altruism (Fehr and Schmidt 1999, Bolton and Ockenfels 2000, Charness and Rabin 2002) and intention-based models of reciprocity (Rabin 1993, Charness and Rabin 2002, Dufwenberg and Kirchsteiger 2004, Falk and Fischbacher 2006, Cox et al. 2008).

As for the former, in our experiment we elicit sensitivity to advantageous inequality (see Game 3 in Section 3.2). However, note that, given the assumption of symmetric cost functions, *i.e.*, $k_i = k_-i$ in all treatments, the equilibrium predictions of Table 1 still hold if players are inequity-averse, since the symmetric strategy $(\beta^*_{bad}, x^*_{bad})$ in the High-cost treatment and the symmetric strategies $(\beta^*_{bad}, x^*_{bad})$, $(\beta^*_{int}, x^*_{int})$, or $(\beta^*_{good}, x^*_{good})$ in the Low-cost treatment also minimize advantageous and disadvantageous inequality: they are both zero in all these symmetric equilibria. This leads to elaborate the following hypothesis.

Behavioral Hypothesis 5 [Independence from inequity aversion]: The levels of investment and effort are independent from the player's sensitivity to inequity aversion. This holds in each treatment.

[9]Note that this hypothesis implicitly assumes no interplay between players' risk attitudes and the treatment variables in our finitely repeated setting. This contrasts with experimental evidence in almost-infinitely repeated games with no strategic incentives to cooperate in the constituent game – *e.g.*, the Prisoner's Dilemma –, where it has been shown that pairs with higher (resp., lower) degree of risk aversion cooperate less (resp., more) (see Sabater-Grande and Georgantzís 2002). We rather claim that the presence of a strategic incentive to cooperate in our constituent game – complementary-building investment – plays a first-order effect over the interplay between risk attitude and strategic reputation building in the finitely repeated interaction.

As for intention-based models of reciprocity, although we do not elicit players' sensitivity to reciprocity in our experiment, it is possible to formulate testable hypotheses under the operational assumption of a similar distribution of players' reciprocal types across the four treatments of the experiment.

Observation 1. Intention-based reciprocity in the one-shot game (according to, *e.g.*, Dufwenberg and Kirchsteiger 2004) would lead to multiple equilibria. In the High-cost treatment, for high enough sensitivity to reciprocity, besides the bad equilibrium with no investment, another equilibrium in pure strategies with positive investment would emerge, and a mixed one. In the Low-cost treatment, besides the three aforementioned equilibria, other equilibria in pure and mixed strategies would emerge. This would ultimately make less sharp the comparison between the High-cost and the Low-cost treatment under Stranger matching.

Observation 2. High-reciprocity types in the Partner matching, once having built reputation in the first periods of the repeated game, would maintain it until the last periods, since this reputation is not "strategic" but rather due to sequential positive reciprocity.[10] Therefore, no end-game effect should be observed.

According to Observation 1, in the last periods of the Stranger treatments – where learning dynamics can justify equilibrium predictions – we should find no significant difference between the Low-cost and High-cost treatments in terms of both investment and effort. This would contradict the first part of Behavioral Hypothesis 2.

Considering Observations 1 and 2 together, in the last periods of the Partner treatments – due to genuine altruism counterbalancing the end-game effect – we should find no significant difference between the Low-cost and High-cost treatment in terms of both investment and effort. This would contradict the second part of Behavioral Hypothesis 2.

Finally, considering Observation 2, given that reputation building is not strategic in the sense of our model, we should find no significant difference between the Partner and the Stranger treatment independently from k, the concavity of the investment cost function (Low-cost and High-cost). This would contradict the first part of Behavioral Hypothesis 3.

Therefore, verification of Behavioral Hypotheses 2 and 3 would not only provide a direct experimental test to our model, but also indirectly show that in our complementarity-building investment game altruistic behavior is not due to exogenous other-regarding preferences. Also selfish players may behave altruistically, if the strategic features of the game are able to endogenously induce strategic other-regarding ethics.

3 Experimental Design

3.1 Experimental procedures

The experiment was programmed using the web platform EconPlay and was run in September-October 2013 on a computer network with 80 inexperienced students at the BETA Laboratory of Experimental Economics (LEES) at the University of Strasbourg.

[10] For the positive interplay between belief-dependent motivations – as guilt and reciprocity – and reputation building in repeated social dilemma games, see Attanasi et al. (2018a).

As for participants' features: 56% male vs. 44% female, average age 22 (min 18, max 31), 74% undergraduate vs. 36% graduate students of different fields: 49% in Human and Social Sciences, 14% in Hard Sciences, 10% in Natural Sciences, 3% in Arts, and 24% in other fields.

Four sessions were organized, with 20 participants per session.[11] Each subject only participated in one session (between-subject design). At the beginning of the session, each participant was randomly assigned to a computer terminal, being physically isolated from the other terminals.

During each session, subjects participated in three games. Participants were given written instructions on the rules of game 1. Instructions for game 2 were given only after game 1 was completed, and instructions for game 3 were given only after game 2 was completed. Draws and payments for each game were made at the end of the experiment, after game 3 was completed.

Participants were paid the sum of their earnings in the three games. Average earnings were 20.2 euro (minimum and maximum earnings were respectively 7 euro and 62.5 euro); the average duration of a session was 1h50 minutes, including instructions and payment.

3.2 Experimental games

The experiment consists of three games.[12]

Game 1 is the implementation of the complementarity-building investment-effort game analyzed in Section 2, according to the parametrization in eq. (12) and the four treatments introduced in Section 2.3. Treatments are implemented according to a between-subject design.

Games 2 and 3 are the same in each treatment. They are meant to elicit respectively the participants' degrees of risk aversion and inequity aversion, needed to test the independence of players' investments and efforts from their sensitivity to risk and inequity aversion (Behavioral Hypotheses 4 and 5, respectively).

Game 1. In both the Stranger and the Partner matching, the 25 periods of the repeated two-stage game are divided in 5 sequences of 5 periods each. At the beginning of each sequence (*i.e.*, $t = 1, t = 6, t = 11, t = 16, t = 21$), 10 pairs of subjects are formed. Within each pair, investments β_1 and β_2 are chosen simultaneously, made public information in the pair, and kept constant for the 5 periods of the sequence. Then, in each of the five periods of the sequence, efforts x_1 and x_2 are simultaneously chosen, and uniperiodal profits are revealed within the pair at the end of each period. Therefore, as in several experimental studies in repeated two-stage games with a long-term variable in the first stage and a short-term variable in the second stage,[13] we maintain fixed for a given number of periods the long-run variable (investment) and we let the short-run variable (effort) change in each period. Among other advantages (reduction of choice overload, less interaction between the two variables, etc.) this should increase subjects'

[11] Two sessions were run on September 30, 2013, and the other two on October 1, 2013.

[12] For the experimental instructions, see the Online Appendix at www.giuseppeattanasi.wixsite.com/index/working-papers.

[13] See, *e.g.*, Barreda-Tarrazona et al. (2011, 2016), Camacho-Cuena et al. (2005), Fatás et al. (2005, 2013), and Vasileiou and Georgantzs (2015).

attention in the choice of the main variable under investigation (the complementarity-building investment), and let them use the effort choice for uniperiodal adjustments to the co-players' fix investment and variable efforts within a sequence. Note that the equilibrium predictions in Table 1 still apply to this version of the two-stage investment-effort game.

At the beginning of each sequence and in each of the five periods of the sequence, before making their choice, subjects have the option to use a gain simulator. For each combination of the four inputs β_1, x_1, β_2 and x_2, *i.e.*, a subject's choices and beliefs, the gain simulator provides as output his/her own uniperiodal gain. Each subject is allowed to use the simulator as many times as he/she likes both at the beginning of the sequence – when choosing β_i – and in each period of the sequence – with β_1 and β_2 chosen at the beginning of the sequence. In the data analysis, we will use the number of times a subject has used the simulator as a proxy for his/her understanding of the strategic features of the game.

In each of the two Stranger treatments, subjects are re-matched at the beginning of each period according to an absolute stranger matching protocol. Then, when choosing β_i at the beginning of each sequence, subjects are aware that in each period of the sequence they will be matched with a different β_{-i} chosen at the beginning of the same sequence. Then, in each period of the sequence, they choose x_i after having observed the β_{-i} of the co-player they have been randomly matched with in that period.

In each of the two Partner treatments, pairs are fixed for all 25 periods, with subjects in the same pair choosing β_1 and β_2 at the beginning of each sequence knowing that they will always play with the same subject (the same as in all other four sequences) in each period of the sequence, where β_{-i} will always be the same. Therefore, in each period of a sequence, they choose x_i after having observed the (same) β_{-i} of the (same) co-player.

Independently of the treatment, only one over the 25 periods of Game 1 is randomly selected for payment at the end of the experiment.

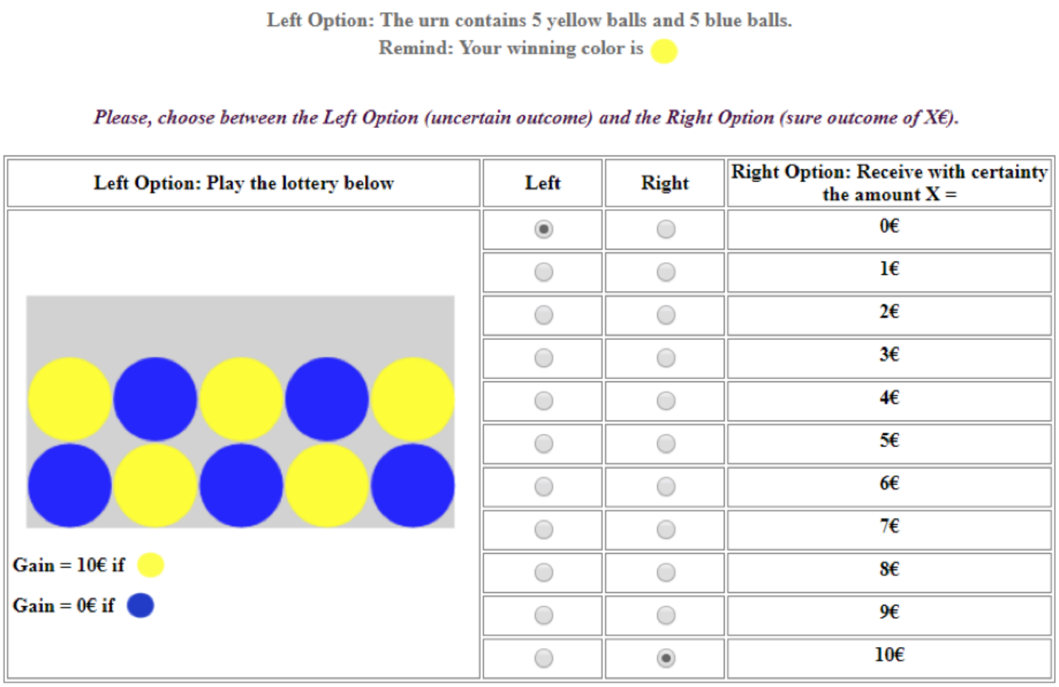
Game 2. It uses Chakravarty and Roy's (2009) method to elicit the participants degree of risk aversion (see Figure 3).[14]

Each participant makes a choice between two options in a list of prizes, similar to the one proposed by Holt and Laury (2002). The left-hand option gives for each line an equal probability of receiving 10 euro or 0 euro (urn of known composition: 5 blue and 5 yellow balls). The right-hand option gives a sure amount that varies from 0 euro (line 1) to 10 euro (line 11). We exogenously impose monotonic behavior, so given the line at which a participant switches from the left-hand option to the right-hand option, the computer automatically assigns the right-hand option to all following lines. At the end of the experiment, the computer randomly draws one of the eleven lines. If the player selected the right-hand option for the drawn line, he or she receives the corresponding sure amount, and if the player selected the left-hand option, he or she receives 10 euro or 0 euro depending on the randomly drawn ball.

As in Chakravarty and Roy (2009) and follow-up studies, we make the following auxiliary assumption: the lower the line number at which a subject selects the right-hand

[14]For previous experimental studies using the same Econplay software for Game 2, see d'Albis et al. (2018).

Figure 3: Experimental Game 2 (multiple price-list method).



option, the higher his or her elicited degree of risk aversion under the expected utility framework.[15]

Game 3. It is the "advantageous inequality-aversion game," which elicits the participants' sensitivity to aversion to advantageous inequality, according to the Fehr and Schmidt (1999) model. It is a modified version of the dictator game proposed by Blanco et al. (2011), first implemented by Attanasi et al. (2019a).

In the game, each subject is matched with another player in the room. Then each participant goes through a list and makes a choice between two options for each line in the list. The left-hand option gives 10 euro to player A and 0 euro to player B, while the right-hand option gives the same amount to the two players, from a minimum of 0 (line 1) to a maximum of 10 (line 11). Each subject knows that his or her decision is relevant only if he or she is finally selected as player A. As in game 2, when the participant selects the right-hand option for a given line, the computer automatically assigns the right-hand option for all the following lines (exogenously-imposed monotonicity). At the end of the experiment, the computer randomly selects within each pair who is player A and who is player B, and randomly draws one of the eleven lines. If a participant is selected as player A, his/her payoff will depend on his/her selected option for the drawn line, while if he/she

[15]Note that this assumption is less general than it might seem at first glance: Attanasi et al. (2018b) have shown that in elicitation tasks à la Holt and Laury (2002) this assumption also holds for non-expected-utility maximizers.

Figure 4: Experimental Game 3 (advantageous inequality-aversion game).

Left Option: 10€ for player A and 0€ for player B	Left	Right	Right Option: X€ for player A and X€ for player B
(10€,0€)	◉	○	(0€,0€)
(10€,0€)	○	○	(1€,1€)
(10€,0€)	○	○	(2€,2€)
(10€,0€)	○	○	(3€,3€)
(10€,0€)	○	○	(4€,4€)
(10€,0€)	○	○	(5€,5€)
(10€,0€)	○	○	(6€,6€)
(10€,0€)	○	○	(7€,7€)
(10€,0€)	○	○	(8€,8€)
(10€,0€)	○	○	(9€,9€)
(10€,0€)	○	◉	(10€,10€)

is selected as player B, his/her payoff will depend on his/her partners selected option.

The line number at which a participant selects the right-hand option gives a rough estimate of his/her sensitivity to aversion to advantageous inequality. This in turn represents a lower bound for his or her sensitivity to aversion to disadvantageous inequality. Given the correlation between the two sensitivities assumed at the individual level by Fehr and Schmidt (1999) model and empirically detected in follow-up experimental studies in dictator games (see, *e.g.*, Saucet and Villeval 2018), we make the following auxiliary assumption: the lower the line number at which a subject selects the right-hand option, the higher his or her elicited sensitivity to aversion to advantageous and disadvantageous inequality.

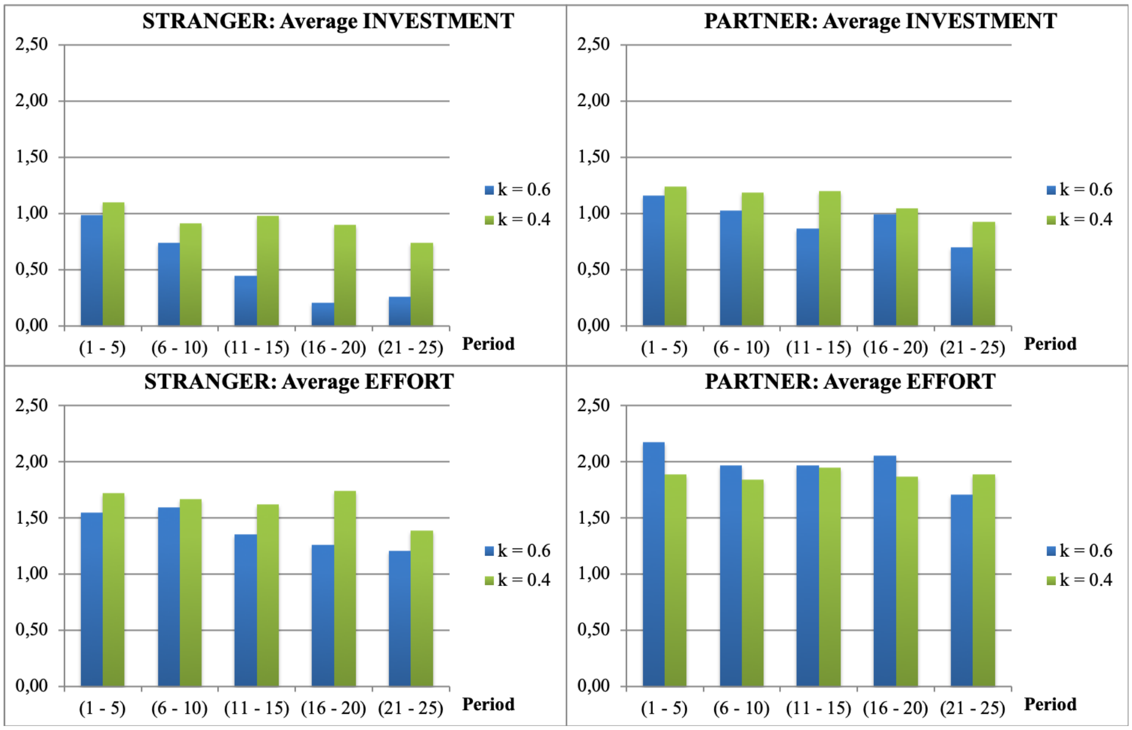
4 Results

4.1 Investment and effort across periods

Figure 5 reports, for each of the five sequences of the repeated game, the subjects' average investment β (top panels) and average effort x (across the five periods of each sequence, bottom panels), for the four treatments. In particular, the left panels concern the Stranger treatments, and the right panels the Partner treatments. Each of the four panel presents subjects' behavior disentangled by the concavity k of investment cost function (High-cost $k = 0.6$ vs. Low-cost $k = 0.4$).

First of all, compare histograms in top and bottom panels of Figure 5, by keeping color (*e.g.*, blue) and side of the panel (*e.g.*, left) fixed. Note that, in line with Behavioral Hypothesis 1, the subject's effort is positively correlated with his/her own investment (Spearman's $\rho = 0.38$, *p-value* $= 0.000$) and especially with the co-player's investment (Spearman's $\rho = 0.68$, *p-value* $= 0.000$). This holds independently of the treatment:

Figure 5: Average Investment and Effort, by sequence and treatment.



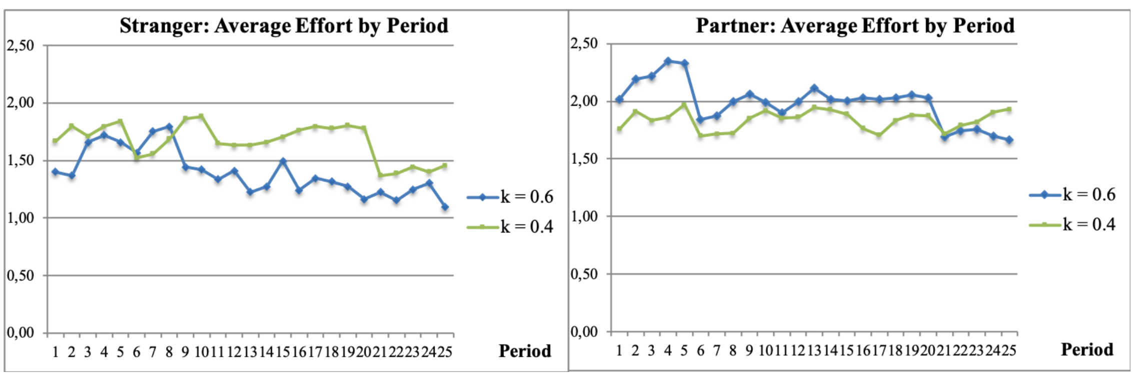
smallest $\rho = 0.14$ (*p-value* $= 0.002$) for investment, smallest $\rho = 0.53$ (*p-value* $= 0.000$), both in the High-cost Stranger treatment. Therefore, we can conclude that **Behavioral Hypothesis 1 is supported**.

Let us now test Behavioral Hypothesis 2 [High- vs. Low-cost] for the investment. First, as the top-left panel of Figure 5 shows, in the Stranger treatments, the investment is significantly higher in the Low-cost than in the High-cost treatment (0.93 vs. 0.53), as it is confirmed by a Mann-Whitney test with a pairwise comparison between treatments (*p-value* $= 0.000$). This also holds if we only consider the last sequence of the repeated game (periods 21-25 in Figure 5): average investment $= 0.74$ in the Low-cost vs. 0.26 in the High-cost, *p-value* $= 0.000$, Mann-Whitney test. Second, as the top-right panel of Figure 5 shows, in the Partner treatments, the investment is higher in the Low-cost than in the High-cost treatment (1.12 vs. 0.95), but the difference is not significant (*p-value* $= 0.245$). This also holds if we only consider the first sequence of the repeated game (periods 21-25 in Figure 5): average investment $= 1.24$ in the Low-cost vs. 1.16 in the High-cost. The treatment difference is lower than by considering all the 25 periods of the game (0.08 vs. 0.18), and still not significant (*p-value* $= 0.297$). Therefore, we can conclude that **Behavioral Hypothesis 2 is supported for the investment**.

Let us now test Behavioral Hypothesis 3 [Stranger vs. Partner matching] for the investment. First, compare the blue histograms in the two top panels of Figure 5 (High-cost treatments). Across all the five sequences of periods, the investment is significantly higher in the Partner than in the Stranger treatment (0.95 vs. 0.53, on average over the five sequences, *p-value* = 0.000). This is also true if we disentangle by sequence (highest *p-value* = 0.034 in the first sequence, periods 1-5). Then, compare the green histograms in the two top panels of Figure 5 (Low-cost treatments). Across all the five sequences of periods, the investment is significantly higher in the Partner than in the Stranger treatment (1.12 vs. 0.93, on average over the five sequences, *p-value* = 0.000), although the difference is smaller than in the High-cost treatments (0.19 vs. 0.42). And in fact, if we disentangle by sequence, the difference is significant only in the second (periods 6-10, *p-value* = 0.006) and the third sequence (periods 11-15, *p-value* = 0.029). In particular, it is not significant in the first sequence (periods 1-5, *p-value* = 0.217). With this, we can state that **Behavioral Hypothesis 3 is supported for the investment**.

We now test Behavioral Hypotheses 2 and 3 for the effort. Figure 6 extends the two bottom panels of Figure 5 by reporting subjects' average effort x in each of the 25 periods of the repeated game, for the four treatments. As for Figure 5, the left panel concerns the two Stranger treatments, and the right panel the two Partner treatments, with the same color code (blue for the High-cost and green for the Low-cost treatment).

Figure 6: Average Effort, by period and treatment.



As for Behavioral Hypothesis 2, consider first the left panel of Figure 6: in the Stranger treatments, the effort is significantly higher in the Low-cost than in the High-cost treatment (1.63 vs. 1.39, *p-value* = 0.000, Mann-Whitney test). The opposite occurs in the Partner treatments, with the effort being significantly lower in the Low-cost than in the High-cost treatment (1.98 vs. 1.89, *p-value* = 0.007, Mann-Whitney test). Here our prediction was of a less significant difference with respect to the Stranger treatments, although still in favor of the Low-cost treatment. However, the direction of the treatment variation is the right one – with the positive effect of strategic complementarity of a lower investment cost being offset by reputation building (Partner matching). And in fact this especially occurs in the first five periods (first sequence) of the repeated game – as the second part of Behavioral Hypothesis 2 suggests – where the positive effect of reputation building is strongest (average effort = 2.17 in the High-cost vs. 1.89 in the Low-cost treatment, *p-value* = 0.003). In confirmation of this, and this time fully in line with our predictions, in the last five periods of the repeated game – where the effect of reputation building is smallest – the average level of effort reverts to be higher in the Low-cost (1.89) than in the High-cost treatment (1.71), with the difference being non-significant (*p-value* = 0.773) and, more importantly, lower than in the Stranger treatments. All this leads us to conclude that **as for the effort, the first part of Behavioral Hypothesis 2 is supported, and the second part only in the last sequence of the repeated game**.

Finally, as for Behavioral Hypothesis 3, given the investment cost function, we find that across all periods the level of effort is significantly higher in the Partner than in the Stranger treatment (*p-value* = 0.000 both in the High-Cost – 1.98 vs. 1.34 – and in the Low-Cost case – 1.89 vs. 1.63). However, restricting the analysis to the first periods of the repeated game, the difference in reputation building is still significant in the High-cost treatments (2.17 vs. 1.55, *p-value* = 0.000) but not in the Low-cost treatments (1.89 vs. 1.73, *p-value* = 0.105). With this, we can state that **Behavioral Hypothesis 3 is supported for the effort**.

4.2 Determinants of the investment in the first stage

To provide econometric support to the results shown in the top panels of Figure 5, we perform parametric regressions to test the effect of the cost of investment k (High vs. Low), the matching protocol (Partner vs. Stranger), and the elicited sensitivities to risk and inequity aversion on the level of investment β, by controlling for relevant covariates. We consider six main regressors. The first four concern Game 1. They are the two treatment dummies of Game 1 – Low Cost (taking value 1 in treatments with $k = 0.4$ and value 0 in those with $k = 0.6$), and Partner (taking value 1 in the Partner treatments and value 0 in the Stranger treatments) –, and another dummy capturing the interaction between the two treatment variables, Low-Cost * Partner (taking value 1 in treatment Low-Cost - Partner matching, and value 0 in the other three treatments)

We also consider the number of times a subject has used the gain simulator during Game 1, at the beginning of the sequence before choosing his investment β (No. of Simulations β) and in each period of the sequence before choosing his effort x (No. of Simulations x). As one might expect, the two variables are positively correlated: subjects making more simulations at the beginning of a sequence are also making more simulations in each period of the sequence (Spearman's $\rho = 0.28$, *p-value* = 0.000). Therefore, we only include in the regression the second variable, No. of Simulations x, because of a higher dispersion, by construction (subjects can use the simulator for β only at the beginning of each of the 5 sequences, while they can use the simulator for x given β in each of the 25 periods of the repeated game).

The other two main regressors are the subject's degree of risk aversion elicited through Game 2, and the subject's sensitivity to inequity aversion elicited through Game 3. The former takes integer values from 1 to 10: it equals 10 if in Game 2 the subject switches to the sure amount at $X = 1$ (first line), ..., and 1 if he/she switches to the sure amount at $X = 10$ (last line). The latter also takes integer value from 1 to 10: it equals 10 if in Game 3 the subject switches to the symmetric payoff profile at $(1, 1)$ (first line), ..., and 1 if he/she switches to the symmetric payoff profile at $(10, 10)$ (last line).

Finally, we consider controls for individual characteristics: gender, age, education level, Human and Social Sciences student dummy (capturing 49% of the subject pool, Economics students being a sub-sample of this category). Education level is not included in the regressions since it positively correlates with age (Spearman's $\rho = 0.57$, *p-value* $= 0.000$).

We run four regressions. The first two regressions (Model I) rely on the fact that data have been obtained through a laboratory experimental setting, where individuals are randomly selected. Since, independently of the matching protocol, β is chosen five times – *i.e.*, at the beginning of each sequence – we treat each sequence of five periods as an independent observation. Hence, we cluster standard errors by sequence. In Model I.a we only have the main regressors, while in Model I.b we also control for individuals' characteristics (gender, age, education level, Human and Social Sciences student dummy).

As a robustness check, we perform other two regressions with random effect estimations controlling for sequence fixed effects. In Model II.a we only have the main regressors, while in Model II.b we also control for individuals' characteristics.

Results of the regression analysis are shown in Table 2. Since $\beta \in [0, 3]$, one might think that Tobit models are the right option. However, we use OLS and GLS models for the four regressions in Table 2 (for Models I and II, respectively), since, in at least one treatment, both 0 and 3 are equilibrium predictions of our theoretical model of Section 2. However, as a robustness check, we have also run the four regression with Tobit models, left censored at 0, and right censored at 3, and found that results are not affected (results are available upon request).

First of all, results in Table 2 provide further support to the first part of Behavioral Hypotheses 2 and 3 regarding the investment. In fact, the coefficients of both the Low-cost and the Partner dummy are always positive, independently of the specification of the regression model and of whether we include subjects' idiosyncratic features in the regression. The Low-cost dummy is significant in three over four model specifications, and almost significant in the last one (*p-value* $= 0.144$). The Partner dummy is significant in all four model specifications.

The negative sign of the coefficient of the interaction term Low*Partner – although not significant in any of the four specifications – shows that the Low-cost vs. High-cost difference in the investment level is less significant in the Partner treatments (second part of Behavioral Hypothesis 2) and that the Partner vs. Stranger difference is less significant in the Low-cost treatments (second part of Behavioral Hypothesis 3).

As for the number of times a subject has used the gain simulator in the 25 periods (No. of Simulations x), recall that we consider it as a proxy for his/her understanding of the strategic features of the game. The significant positive (although, very low) coefficient of "No. of Simulations x" lets us state that subjects who had better understood the complementarity-building game have invested more in complementarity-building within the pair. This seems to suggest that a subject's higher level of rationality makes him/her get and exploit the strategic role of complementarity-building investment. The latter is the main ingredient of our theoretical model and of the equilibrium predictions.

The analysis of sign, size and significance of the coefficients of the variables Risk Aversion and Inequity Aversion in Table 2 helps us verify Behavioral Hypotheses 4 and 5, respectively. Note that for both variables, the coefficient is negative but very close to 0, and never significant. This holds independently of the regression model (smallest

Table 2: Marginal effects from ordinary least squares regressions with clustered standard errors per session (Model I), and with random effects (Model II), explaining investment β

Dep. variable: β	Model I.a	Model I.b	Model II.a	Model II.b
Low-Cost	0.358 **	0.292 *	0.359 *	0.293
	(0.101)	(0.131)	(0.191)	(0.201)
Partner	0.413 **	0.339 *	0.414 **	0.340 *
	(0.102)	(0.126)	(0.186)	(0.197)
Low-Cost * Partner	-0.152	-0.105	-0.155	-0.109
	(0.113)	(0.138)	(0.266)	(0.276)
Risk Aversion	-0.024	-0.032	-0.024	-0.032
	(0.017)	(0.017)	(0.037)	(0.037)
Inequity Aversion	-0.018	-0.020	-0.018	-0.020
	(0.010)	(0.012)	(0.032)	(0.034)
No. of Simulations x	0.009 *	0.009 *	0.008 ***	0.008 ***
	(0.004)	(0.004)	(0.002)	(0.002)
Female	-	0.091	-	0.091
	-	(0.076)	-	(0.143)
Age	-	-0.035 *	-	-0.035
	-	(0.013)	-	(0.027)
Human and Social Sciences	-	-0.161 ***	-	-0.160
	-	(0.029)	-	(0.139)
Constant	0.257	1.069	0.496 **	1.311 *
	(0.219)	(0.566)	(0.252)	(0.672)
R-squared	0.079	0.097	0.111	0.130
No. of observations	2000	2000	2000	2000

Note: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

p-value $= 0.140$ for Risk Aversion and $= 0.157$ for Inequity Aversion), and it is confirmed by non-parametric tests: rank correlation between Risk Aversion and investment is low although significant at the 5% level (Spearman's $\rho = -0.10$, *p-value* $= 0.049$), and between Inequity Aversion and investment is low and non-significant (Spearman's $\rho = -0.08$, *p-value* $= 0.102$).

Disentangling by treatment, Risk Aversion and investment are significantly and negatively correlated only in the High-Cost Stranger treatment (Spearman's $\rho = -0.19$, *p-value* $= 0.058$), while Inequity Aversion and investment are significantly correlated in all treatments but the High-Cost Stranger treatment. However, in the two Low-Cost treatments we detect a negative correlation (Spearman's $\rho = -0.21$, *p-value* $= 0.040$ in the Low-Cost Stranger, Spearman's $\rho = -0.31$, *p-value* $= 0.002$ in the Low-Cost Partner), while the correlation is positive in the High-Cost Partner treatment Spearman's $\rho = 0.20$, *p-value* $= 0.045$).

Therefore, we can conclude that **Behavioral Hypotheses 4 is supported for the investment in all treatments apart from the High-Cost Stranger one**. In the latter treatment risk aversion negatively correlates with complementarity-building investment.

Conversely, **Behavioral Hypothesis 5 is supported for the investment only in the High-Cost Stranger treatment**. However, in the other treatments inequity aversion does not seem to play a univocal role on the investment, the sign of the correlation being negative in the Low-Cost treatments, and positive in the High-Cost Partner treatment.

As for subjects' idiosyncratic features, we see that both age and the fact of being a student in Human and Social Sciences have a negative impact on investment. However, none of the two coefficients is significant when we consider random effects (Model II.b).[16] Finally, we find no gender effect.

4.3 Determinants of the effort in the second stage

Table 3 provides econometric support to the results shown in the bottom panels of Figure 5 and in Figure 6. As we did for investment in Table 2, we perform parametric regressions to test the effect of the cost of investment k (High vs. Low), the matching protocol (Partner vs. Stranger) and the elicited sensitivities to risk and inequity aversion on the level of effort x, by controlling for relevant covariates.

Besides all the regressors of Table 2, we consider other two explanatory variables, needed to provide further test to Behavioral Hypothesis 1. In fact, recall that both in the Stranger and in the Partner treatments, in each period subjects choose the effort after having chosen at the beginning of the sequence their own investment, and having observed the investment chosen by their co-player at the beginning of the same sequence. Variables Own Investment and Other Investment respectively represent these two choices.

In the first two OLS models, since the effort is chosen in each of the 25 periods of the repeated game, we cluster standard errors by period, by also controlling for individuals' characteristics in Model I.b. As for the investment in Table 2, for robustness check we perform other two regressions, GLS models with random effect estimations controlling for period fixed effects, by also controlling for individuals characteristics in Model II.b.

First of all, results in Table 3 provide further and strong support to Behavioral Hypothesis 1. In fact, the coefficients of both Own Investment and Other Investment are always positive and highly significant, independently of the specification of the regression model and of whether we include subjects' idiosyncratic features in the regression. Note that the coefficient of Other Investment is always more than twice than the one of Own Investment. This shows the strategic complementarity-building effect of a player's investment in the first stage on the co-player's effort on the second stage, which is indeed the main ingredient of our model.

Results in Table 3 provide further support to the first part of Behavioral Hypothesis 3 regarding effort. In fact, the coefficient of the Partner dummy is always positive and highly

[16] As for the Human and Social Sciences dummy, which is highly significant in Model I, recall that in the complementary fields of study around half of the sample (24%) are students in Hard Sciences, or in Natural Sciences, which are supposed to have a better background in Mathematics than students in Human and Social Sciences. And in fact the latter use the gain calculator significantly more than the former (around 8 times vs. 6 times per period).

Table 3: Marginal effects from ordinary least squares regressions with clustered standard errors per period (Model I), and with random effects (Model II), explaining effort x

Dep. variable: x	Model I.a	Model I.b	Model II.a	Model II.b
Own Investment	0.188 ***	0.192 ***	0.143 ***	0.144 ***
	(0.013)	(0.013)	(0.018)	(0.018)
Other Investment	0.479 ***	0.481 ***	0.427 ***	0.427 ***
	(0.023)	(0.023)	(0.018)	(0.018)
Low-Cost	-0.015	-0.008	0.023	0.026
	(0.036)	(0.038)	(0.070)	(0.073)
Partner	0.304 ***	0.309 ***	0.346 ***	0.345 ***
	(0.036)	(0.042)	(0.068)	(0.071)
Low-Cost * Partner	-0.208 ***	-0.208 ***	-0.233 **	-0.229 **
	(0.054)	(0.058)	(0.097)	(0.100)
Risk Aversion	0.007	0.008	0.006	0.006
	(0.006)	(0.006)	(0.013))	(0.013)
Inequity Aversion	0.007 *	0.003	0.005	0.001
	(0.004)	(0.004)	(0.012)	(0.012)
No. of Simulations x	-0.005 *	-0.005 **	-0.006 ***	-0.006 ***
	(0.002)	(0.002)	(0.002)	(0.002)
Female	-	0.020	-	0.036
	-	(0.021)	-	(0.052)
Age	-	0.006	-	0.003
	-	(0.005)	-	(0.010)
Human and Social Sciences	-	0.059***	-	0.042
	-	(0.015)	-	(0.050)
Constant	1.146 ***	0.951***	1.046 ***	0.922 ***
	(0.040)	(0.102)	(0.108)	(0.251)
R-squared	0.485	0.487	0.489	0.491
No. of observations	2000	2000	2000	2000

Note: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

significant, independently of the specification of the regression model and of whether we include subjects' idiosyncratic features in the regression.

However, the first part of Behavioral Hypothesis 2 is not supported for the effort by any of the regression models of Table 3. In fact, the coefficient of the Low-Cost dummy is almost null and never significant. This can be explained by the higher level of effort in the first four sequences of the High-Cost than the Low-Cost treatment under partner matching, as previously shown in the left panel of Figure 6. As a matter of fact, the coefficient of the interaction term Low*Partner is negative and significant in each of the four specifications. Therefore, the positive interaction between High-Cost and Partner matching in reputation building counterbalances the strategic effect of low complementarity-building investments under Stranger matching. This ultimately leads to a non-significant effect of the Low-Cost dummy on effort when considering the four treatments as a whole. Recall, however, that variables Own Investment and Other Investment in Table 3 already capture the positive effect of the Low-Cost dummy on investment shown in Table 2. Therefore, although the direct effect of Low-Cost on effort is null, the indirect effect – mediated by the investment – is positive. This provides some support to Behavioral Hypothesis 2.

The number of times a subject has used the gain simulator in the 25 periods (No. of Simulations x), contrarily to what we find for the investment, has a negative impact on effort, although – as for the investment – very low. This might lead to conclude that subjects who had better understood the complementarity-building game invest slightly more in complementarity-building within the pair, but then exert slightly less effort to compensate, eventually because the former is high from reputation building in a repeated setting like ours.

As for Risk Aversion and Inequity Aversion, Table 3 shows that for both variables the coefficient is positive but very close to 0 independently of the regression model (recall that it was negative and very close to 0 for the investment in Table 2). Furthermore, Risk Aversion is never significant, independently of the regression model (smallest *p-value* = 0.226 in Model I.b). This is confirmed by a Spearman's test of rank correlation between Risk Aversion and effort: $\rho = -0.002$, *p-value* = 0.935. On the same line, Inequity Aversion is significant only in one of the four models and only at the 10% level, and a Spearman's test of rank correlation confirms absence of interaction with the effort: $\rho = -0.03$, *p-value* = 0.167.

Disentangling by treatment, Risk Aversion and effort are significantly correlated in only two of the four treatments. Furthermore, the correlation is low in both treatments, positive in the Low-Cost Stranger treatment (Spearman's $\rho = 0.16$, *p-value* = 0.000), and negative in the High-Cost Partner treatment (Spearman's $\rho = -0.14$, *p-value* = 0.002). Therefore, we can conclude that **Behavioral Hypotheses 4 is supported for the effort only in the High-Cost Stranger and the Low-Cost Partner treatment**.

The picture of the interaction between Inequity Aversion and effort across treatments is similar. They are significantly correlated in only two of the four treatments, with this correlation being low and positive in the Low-Cost Partner treatment (Spearman's $\rho = 0.09$, *p-value* = 0.000), and negative and low in the High-Cost Partner treatment (Spearman's $\rho = -0.16$, *p-value* = 0.034). Therefore, we can conclude that **Behavioral Hypothesis 5 is supported for the effort only in the High-Cost Stranger treatment**.

As for subjects' idiosyncratic features, we find neither gender nor age effect. Only the fact of being a student in Human and Social Sciences has a significant impact on effort, which is negative (it was positive on the investment). But this occurs only in Model I.b, as the coefficient becomes non-significant when we consider random effects (Model II.b).

5 Concluding remarks and extensions

Reciprocity may arise in the absence of any reputation effect or cooperation due to repeated play. In this paper, we have shown theoretically and experimentally how strategically reciprocal behavior may emerge from a rational individual's interest to increase another's capacity to benefit from his own actions. The one-shot strategic situation we analyze is a two-stage game where in the first stage two players simultaneously choose their own level of complementarity-building investment and in the second stage they simultaneously choose their own level of effort. Players' benefits from own effort are enhanced by both the co-player's investment and the co-player's effort, but not by their own investment.

In this context, we have first proven theoretically that an agent will strategically incur a cost (investment cost), without directly benefiting from it, in order to help the other benefit from synergies arising due to both agents' efforts. Therefore, what is typically perceived as an altruistic behavior may instead be due to a purely strategic effect, which in turn leads to an increase of the two players' interdependence in the effort-choice stage of the game.

Our theoretical analysis predicts that if human interaction is the result of situations like the ones described here, three types of outcomes should be expected. The good one, in which everyone does everything possible for others to benefit from their own effort, extracting everybody's maximal effort and yielding maximal utility to all. The bad one, in which nobody invests in complementarity building, extracting minimal, individually optimal effort levels and a positive but minimal equilibrium utility resulting from isolated human action. The interior equilibrium in which agents contribute a positive but limited amount in complementarity building, yielding a similar effort level and a higher utility than in the good equilibrium (see Table 1 in Section 2.3). The latter is an intriguing focal point where agents realize that they do not need maximal investment to induce the highest possible level of effort.

The experimental test confirms our theoretical predictions. In particular, we have implemented the complementarity-building investment-effort game under two different specifications of the cost of investing in the other's ability of profiting from synergies. Our model predicts that under the high-cost specification only the bad equilibrium is possible, while under the low-cost specification, all the three abovementioned equilibria exist. Consistently with our main theoretical predictions, we observe that the level of effort optimally increases both with own and other's complementarity-building investment (Behavioral Hypothesis 1), and that the levels of the complementarity-building investment are significantly higher in the presence of a lower cost (Behavioral Hypothesis 2). Furthermore, the strategic complementarity building effect survives the confound of strategic reputation building when the investment-effort game is repeatedly played under partner matching (Behavioral Hypothesis 3). Finally, we find that the strategic complementarity building effect is independent of both psychological attitudes (risk aversion, Behavioral Hypothesis 4) and social attitudes (inequity aversion, Behavioral Hypothesis 5) usually detected in experimental settings similar to ours.

We interpret the confirmation of our theoretical predictions as evidence of the fact that in our complementarity-building investment game altruistic behavior is not due to risk or other-regarding preferences. Also selfish expected-payoff maximizers may behave altruistically, if the strategic features of the game are able to endogenously induce strategic other-regarding ethics.

Surprisingly, in many real-world situations people find it hard to reach the good equilibrium. So far, this has been explained as a consequence of people's lack of altruism or, alternatively, as dynamic instability of the cooperative outcome in situations of repeated interaction. In our framework, failure to reach the good equilibrium can only be due to lack of synergies arising from interacting agents' efforts. Certainly, a more plausible explanation could be that agents often fail to recognize the benefits from actions which benefit others, if such actions do not have a direct positive impact on their own utility. Then, it would be due to strategic myopia and not to lack of altruism that good equilibria may not be reached as often as they should. Fortunately, some cultures seem to realize the benefits from altruistic transmission of aptitudes and habits, thus rendering culture a healthy self-sustained organism.

We conclude by proposing three possible extensions of our analysis, that we leave for further research.

First, in the theoretical analysis and related experimental test of this paper we only focus on symmetric complementarity-building investment games, *i.e.*, with $k_i = k_{-i}$ in eq. (1) of Section 2.1. However, it can be shown that if agents differ in their efficiencies to increase each other's synergy-absorbing capacities ($k_i \neq k_{-i}$), asymmetric equilibria arise. The interior equilibria obtained in the asymmetric case are unfair, in the sense that efficient and, thus, generous agents enjoy less utility than their inefficient counterparts.[17] In a more general model in which fairness considerations are relevant, the fact that interior equilibria are, generally speaking, unfair makes them more difficult to emerge and more vulnerable to opportunistic, non-altruistic behavior. The difficulty of such unfair equilibria to emerge is further enhanced by the fact that they are unstable. Furthermore, if $k_i > 1/2$ and $k_{-i} < 1/2$, then one could find asymmetric situations in which only the aforementioned bad equilibrium exists. The analysis of the asymmetric case would have interesting applications in real-life situations, *e.g.*, to the way in which reciprocal "ethical" altruism may emerge among a senior teacher of a musical instrument and a junior member of the same cultural community.[18]

Second, in several real-life situations complementary-building investments occur sequentially. This is especially true in a context of intergenerational transmission of culture, given that parents, teachers, artists, intellectuals, etc. always play a leader's role when acting as "senders" of cultural transmission messages. Extending our model in this direction is straightforward, with the follower choosing the investment level after having observed the one of the leader, and then leader and follower simultaneously choosing their own level of effort as in the benchmark case. Such analysis may offer a rationale of why altruistic actions are easier to observe when agents act sequentially, within a framework of strategically ethic altruism alternative to models of intergenerational transmission of culture that we have discussed in the Introduction.

Last, it would be interesting to investigate whether explicit cooperation in the effort

[17] For example, if $k_1 = 0.16$ and $k_2 = 0.13$, the resulting interior equilibrium is given by: $(\beta_1^*, \beta_2^*) = (1.0, 1.1)$. In such an equilibrium, the *efficient* agent invests more in the *inefficient* one's absorptive capacity than does the latter in favor of the former. This leads the efficient agent to enjoy lower levels of utility than the inefficient one does, $(1.9, 1.8)$.

[18] Contrarily to formal classical education, folk musicians are usually taught by senior artists without any monetary reward. Of course, the respect that the junior musician may feel for the "maestro" or other signs of family gratitude and gifts may be considered a material reward, but, according to our framework, the true objective of the senior musicians' *investment* is to enable the student understand the local musical terminology, learn the right scales for improvisation, and develop a high degree of understanding for more direct informal communication during a performance (*effort* stage), which usually requires spontaneous arrangements in real time (see, *e.g.*, Attanasi 2007).

stage (like in organized collective activities) facilitates the emergence of the reciprocal equilibrium and whether increasing the group size affects the reciprocal equilibrium described here. Our conjecture is that, if effort synergies enter into an agent's utility as the sum of bilateral interaction terms, agents may (under specific conditions) prefer to specialize in two-person reciprocal relations, whereas if multilateral synergies appear in an agent's utility as the product of many agents' efforts, global (social) reciprocal equilibria are more likely to emerge. In fact, introducing an assumption of decreasing returns to synergy-developing group size could intuitively help define an optimal group size within which complementarity-building investment takes place.

References

[1] d'Albis, H., Attanasi, G. and E. Thibault (2018). An Experimental Test of the Under-Annuitization Puzzle with Smooth Ambiguity and Charitable Giving (No. 18-932). Toulouse School of Economics (TSE).

[2] Ale, S. B., Brown, J. S. and A. T. Sullivan (2013), Evolution of Cooperation: Combining Kin Selection and Reciprocal Altruism into Matrix Games with Social Dilemmas, *PloS one*, 8(5), e63761.

[3] Andreoni, J. (1988), Why Free Ride?: Strategies and Learning in Public Goods Experiments, *Journal of Public Economics*, 37, 291-304.

[4] Attanasi, F. M. (2007), La Musica nel Tarantismo. Le Fonti Storiche, Pisa: ETS.

[5] Attanasi, G. and K. Boun My (2016), Jeu du Dictateur et Jeu de la Confiance: Prfrences Distributives vs Prfrences Dpendantes des Croyances, *L'Actualit Economique*, 92, 249-287.

[6] Attanasi, G., Hopfensitz, A., Lorini, E. and F. Moisan (2016), Social Connectedness Improves Co-ordination on Individually Costly, Efficient Outcomes. *European Economic Review*, 90, 86-106.

[7] Attanasi, G., Dess, R., Moisan, F. and D. Robertson (2017), Public Goods, Role Models and Sucker Aversion: The Audience Matters, CEPR Discussion Papers 12413, C.E.P.R. Discussion Papers.

[8] Attanasi, G., Battigalli, P., Manzoni, E. and R. Nagel (2018a), Belief-dependent Preferences and Reputation: Experimental Analysis of a Repeated Trust Game, *Journal of Economic Behavior and Organization*, in press.

[9] Attanasi, G., Georgantzís, N., Rotondi, V. and D. Vigani (2018b). Lottery-and Survey-based Risk Attitudes Linked through a Multichoice Elicitation Task, *Theory and Decision*, 84, 341-372.

[10] Attanasi, G., My, K. B., Buso, M. and A. Stenger (2019a), Private Investment with Social Benefits under Uncertainty: The Dark Side of Public Financing, *Journal of Public Economic Theory*, in press.

[11] Attanasi, G., Rimbaud, C. and M. C. Villeval (2019b), Embezzlement and Guilt Aversion, *Journal of Economic Behavior and Organization*, in press.

[12] Axelrod, R. (1984), The Evolution of Cooperation, New York: Basic.

[13] Barreda-Tarrazona, I., Garca-Gallego, A., Georgantzís, N., Andaluz-Funcia, J. and A. Gil-Sanz (2011), An Experiment on Spatial Competition with Endogenous Pricing, *International Journal of Industrial Organization*, 29, 74-83.

[14] Battigalli, P. (2019), Game Theory: Analysis of Strategic Thinking, mimeo.

[15] Barreda-Tarrazona, I., Georgantzís, N., Manasakis, C., Mitrokostas, E. and E. Petrakis (2016), Endogenous Managerial Compensation Contracts in Experimental Quantity-setting Duopolies, *Economic Modelling*, 54, 205-217.

[16] Bisin, A. and T. Verdier (2000), Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits, *Quarterly Journal of Economics*, 115, 955-988.

[17] Blanco, M., Engelmann, D. and H. T. Normann (2011), A Withinsubject Analysis of Otherregarding Preferences, *Games and Economic Behavior*, 72, 321-338.

[18] Bolton, G. and A. Ockenfels (2000), A Theory of Equity, Reciprocity and Competition, *American Economic Review*, 90, 166-194.

[19] Camacho-Cuena, E., Garcia-Gallego, A., Georgantzís, N. and G. Sabater-Grande (2005), Buyer-seller Interaction in Experimental Spatial Markets, *Regional Science and Urban Economics*, 35, 89-108.

[20] Chakravarty, S. and J. Roy (2009), Recursive Expected Utility and the Separation of Attitudes towards Risk and Ambiguity: An Experimental Study, *Theory and Decision*, 66, 199-228.

[21] Charness, G. and M. Rabin (2002), Understanding Social Preferences with Simple Tests, *Quarterly Journal of Economics*, 117, 817-869.

[22] Charness, G. and M. C. Villeval (2009), Cooperation and Competition in Intergenerational Experiments in the Field and the Laboratory, *American Economic Review*, 99, 956-978.

[23] Chaudhuri, A., Graziano, S. and P. Maitra (2006), Social Learning and Norms in a Public Goods Experiment with Inter-generational Advice, *Review of Economic Studies*, 73, 357-380.

[24] Cox, J. C., Friedman, D. and V. Sadiraj (2008). Revealed Altruism, *Econometrica*, 76, 31-69.

[25] Della Lena, S., Manzoni, E. and F. Panebianco (2019). On the Transmission of Guilt Aversion, mimeo.

[26] Dufwenberg, M. and G. Kirchsteiger (2004), A Theory of Sequential Reciprocity, *Games and Economic Behavior*, 47, 268-298.

[27] Falk, A. and U. Fischbacher (2006). A Theory of Reciprocity, *Games and Economic Behavior*, 54, 293-315.

[28] Fatás, E., Georgantzís, N., Mánez, J. and G. Sabater-Grande (2005), Pro-competitive Price Beating Guarantees: Experimental Evidence, *Review of Industrial Organization*, 26, 115-136.

[29] Fatás, E., Georgantzís, N., Mánez, J. A. and G. Sabater-Grande (2013), Experimental Duopolies under Price Guarantees, *Applied Economics*, 45, 15-35.

[30] Fehr, E. and K. Schmidt (1999), A Theory of Fairness, Competition, and Cooperation, *Quarterly Journal of Economics*, 114, 817-868.

[31] Gil-Moltó, M. J., N. Georgantzís and V. Orts (2005), Cooperative R&D with Endogenous Technology Differentiation, *Journal of Economics and Management Strategy*, 14, 461-476.

[32] Holt, C. A. and S. K. Laury (2002), Risk Aversion and Incentive Effects, *American Economic Review*, 92, 1644-1655.

[33] Kuang, X. J., Weber, R. A. and J. Dana (2007), How Effective is Advice from Interested Parties?: An Experimental Test using a Pure Coordination Game, *Journal of Economic Behavior and Organization*, 62, 591-604.

[34] Masclet, D. and M. C. Villeval (2008), Punishment, Inequality, and Welfare: A Public Good Experiment, *Social Choice and Welfare*, 31, 475-502.

[35] Masclet, D., Noussair, C. N. and M. C. Villeval (2013), Threat and Punishment in Public Good Experiments, *Economic Inquiry*, 51, 1421-1441.

[36] Masclet, D., Noussair, C., Tucker, S. and M. C. Villeval (2003), Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism, *American Economic Review*, 93, 366-380.

[37] Milliou, C. (2006), Endogenous Protection of RRD Investments, Economics Working WP06-63, Economics Series 25, Universidad Carlos III, Departamento de Economía.

[38] Potters, J. and S. Suetens (2009), Cooperation in Experimental Games of Strategic Complements and Substitutes, *Review of Economic Studies*, 76, 1125-1147.

[39] Rabin, M. (1993), Incorporating Fairness into Game Theory and Economics, *American Economic Review*, 83, 1281-1302.

[40] Sabater-Grande, G. and N. Georgantzís (2002), Accounting for Risk Aversion in Repeated Prisoners' Dilemma Games: An Experimental Test, *Journal of Economic Behavior and Organization*, 48, 37-50.

[41] Saucet, C. and M. C. Villeval (2018), Motivated Memory in Dictator Games, Working Paper GATE 2018-04.

[42] Schotter, A. and B. Sopher (2003), Social Learning and Coordination Conventions in Intergenerational Games: An Experimental Study, *Journal of Political Economy*, 111, 498-529.

[43] Schotter, A. and B. Sopher (2006), Trust and Trustworthiness in Games: An Experimental Study of Intergenerational Advice, *Experimental Economics*, 9, 123-145.

[44] Schotter, A. and B. Sopher (2007), Advice and Behavior in Intergenerational Ultimatum Games: An Experimental Approach, *Games and Economic Behavior*, 58, 365-393.

[45] Smith, A. (1759), *The Theory of Moral Sentiments*, Ed. DD Raphael & AL Macfie, Liberty Fund.

[46] Suetens, S. (2005), Cooperative and Noncooperative R&D in Experimental Duopoly Markets, *International Journal of Industrial Organization*, 23, 63-82.

[47] Trivers, R. L. (1971), The Evolution of Reciprocal Altruism, *Quarterly Review of Biology*, 46, 35-57.

[48] Vasileiou, E. and N. Georgantzís (2015), An Experiment on Energy-saving Competition with Socially Responsible Consumers: Opening the Black Box, *Journal of Behavioral and Experimental Economics*, 58, 1-10.

[49] Villeval, M. C. (2012), Contributions to Public Goods and Social Preferences, *Revue Economique*, 63, 389-420.

[50] Vives, X. (1990), Nash Equilibrium with Strategic Complementarities, *Journal of Mathematical Ecoomics*, 19, 305-321.

[51] Vives, X. (2005), Games with Strategic Complementarities: New Applications to Industrial Organization, *International Journal of Industrial Organization*, 23, 625-637.

[52] Vives, X. (2009), Strategic Complementarity in Multi-Stage Games, *Economic Theory*, 40, 151-171.