We Can Be Heroes

Trust and Resilience in Corrupted Economic Environments[**]

Leonardo Becchetti[*]

Department of Economics and Finance
University of Rome Tor Vergata
Via Columbia, 2 - 00133 Rome
becchetti@economia.uniroma2.it

and School of Finance and Management
SOAS University of London

Luca Corazzini

Department of Economics and Center for Experimental Research in Management and Economics (CERME)
University of Venice "Ca Foscari"
Cannaregio, 821 – 30121 Venice
luca.corazzini@unive.it

Vittorio Pelligra

Department of Economics and Business
University of Cagliari
V. le S. Ignazio, 17 - 09123 Cagliari
pelligra@unica.it

CRENoS

Abstract

We use an original variant of the standard trust game, in order to study the effect of corruption on trust and trustworthiness. In this game, both the trustor and the trustee know that part of the surplus they can generate may be captured by a third "corrupted" player under different expected costs of audit and prosecution. We find slightly higher trustor's giving in presence of corruption, matched by a significant effect of excess reciprocity from the trustee. Both the trustor and the trustee expect on average corruption acting as a tax, inelastic to changes in the risk of corruptor audit. Expectations are correct for the inelasticity assumption, and for the actual value of the "corruption tax". Our experimental findings lead to the rejection of four standard hypotheses based on purely self-regarding preferences. We discuss how the apparently paradoxical excess reciprocity effect is consistent with the cultural role of heroes in history where examples of commendable giving were used to stimulate emulation of the ordinary people. Our results suggest that the excess reciprocity component of the trustee makes trustor's excess giving a rational and effective strategy.

Keywords: trust game, corruption, lab experiment.

JEL Numbers: C72 (Noncooperative games), C91 (Laboratory, Individual behaviour).

[**] We thank Tiziana Medda, Alessio Cordella e Sara Demartis for their invaluable research assistance in running the experiment. Financial support from the University of Rome - Tor Vergata (*Consolidate the Foundations* research grant, P.I. Germana Corrado) and the University of Cagliari (*Fondazione di Sardegna fundamental research grant* L.R. 7/2005, "Economic Growth, Cultural Values and Institutional Design: Theory and Empirical Evidence", P.I. Vittorio Pelligra). All errors are ours.

[*] Corresponding author.

1. Introduction

There is no doubt about the fact that trust is "the lubricant of the social system" to use Arrow's rightly famous expression (1974, p.357). The positive effects associated with being able to trust each other, even in the economic sphere, were already clear to John Stuart Mill, a century earlier, when he noticed that "The advantage of humankind of being able to trust one another, penetrates into every crevice and cranny of human life: the economical is perhaps the smallest part of it, yet even this is incalculable." (1848, p.131). A huge amount of work has been done from then on to deepen our understanding of the role of trust in interpersonal and societal relationships, and to collect empirical evidence on the link between trust and socio-economic relevant variables. We know, for example, that the levels of trust among people are higher in rich nations with low income inequality, in countries with more efficient institutions, and in societies made up by better-educated citizens (Knack and Keefer, 1997). We also know that higher trust has a positive effect on the development of a country's financial system and even to cross-country trade (Guiso *et al.*, 2004, 2009). Trust is also thought as playing a role on people's happiness, a variable difficult to measure, but of increasing importance for the economic discourse. By using a variety of different survey data, Helliwell et al. (2017), find a positive link between trust and the level of subjective well-being: "Whether in the workplace, at home, in the community, or among nations, better and deeper social connections, and especially higher levels of trust are linked to higher subjective well-being, even beyond the effects flowing through higher incomes and better health" (p.1). In a different research, they also find that trust is associated with resilience. In fact, "Being subject to discrimination, ill-health or unemployment, although always damaging to subjective well-being, is much less damaging to those living in trustworthy environments" (Helliwell *et al.*, 2016, p.1).

Being trust such a valuable asset, it is important to come to a deep understanding of its determinants, the factors and the dynamics of its formation and erosion. In this line, with this paper, we try to clarify, by means of a laboratory experiment, the role of corruption, one of the elements that are supposed to affect people's willingness to trust one another and to respond trustworthily to a trusting partner. By corruption, in this context, we intend the possibility that an external (public of private) subject extorts a share of the value produced by another or more subjects. Reasonable examples include a team-leader who takes all the credit for the work done by the team-members, a public official who expects to be paid to facilitate the realization of a public work by a joint venture, as well as a venture capitalist who claims a higher remuneration or a higher degree of control from the startuppers than justified by his or her investment. If entrepreneurial decisions and innovative opportunities depend on "the portion of the value that the venture creates that the entrepreneur is able to capture for their own purposes" (Baker et al., 2005, p. 497), when corruption is present, entrepreneurs and startuppers face a higher risk of opportunistic appropriation of the surplus they are capable to create.

We are interested in investigating to what extent the external costs imposed by corruption, or the act of extortion by one player, affects the level of trust and trustworthiness among the other players.

The economic impact of corruptive practices has inspired, in the recent years, a large body of research that documented, by using a large set of economic indicators including foreign direct investment (Lambsdorff, 2003; Mauro, 1995), productivity (Lambsdorff, 2003; Rivera-Batiz, 2002), income inequality (Li *et al.*, 2000), entrepreneurship (Anokhin and Schulze, 2019) and growth in income (Kaufmann and Kraay, 2003), a negative relationship between the level of corruption and performances. A generally strong and negative correlation has also been found between corruption and trust. However, evidence on the causal link between corruption and trust is at best mixed. On the one hand, in fact, some have taken the view that low levels of trust in a society may favor corruption because of the widespread sense of opportunism (LaPorta et al., 1997; Bjornskov, 2011; Moreno, 2002; Seligson, 2002). Others have shown that a lack of trust may generate the perception of high levels of corruption (Rotondi and Stanca, 2015), that in turn renders corruption more acceptable and likely to occur (Bardhan, 1997; Innes and Mitra, 2013). Many, on the other hand, view corruption as one of the driving forces behind the erosion of trust (Anderson and Tverdova, 2003; Chang and Chu, 2006; Della Porta, 2000). Some have also discussed the possibility of a circular causality (Uslaner, 2008): corrupt officials and business people tend to illegally appropriate an undue share of resources, making, in this way, the rich even richer. So, corruption fuels inequality, which leads to lower trust and even more corruption.

Our experiment addresses the causality problem with respect to the trust-corruption link in a very simplified setting that, however, describes many of the relevant factors. We devise an original variant of the standard investment game (Berg *et al*., 1995) where the trustor and the trustee know that part of the surplus they generate can be extorted by a third party under different regimes of monitoring and subject to different expected audit probabilities.

The only experiment, to our knowledge, that addresses a similar problem is Banerjee's (2016). In his design participants play first either a harassment bribery game, or a strategically equivalent ultimatum game. These games mimic a situation where extortion is an option. In the second stage of the experiment the same players interact in a standard trust game in order to measure the effect of the previous experience on their willingness to trust. Findings from this experiment show a negative spillover effect of corruption on trust.

Our approach is original vis-à-vis the relevant work of Banerjee in several aspects. First of all, we look at the impact of corruption on trust and trustworthiness within the same treatment. This implies that the stylized corruption activity has a direct effect on players' payoffs and does not enter into the game as a spillover effect produced by results and characteristics of a previous and independent treatment. In this respect, our treatment has an important element of external consistency since corruption is modelled as extorting a share of the expected payoff of the trustor and the trustee and as such, affecting, directly, their decisions of trust and reciprocity. Second, the above described approach allows us to measure directly the expected perceived corruption and to analyze how it affects trustors' and trustees' choices. Third, we evaluate the impact of different policies and, specifically, the relative deterrence impact of a high versus low probability of audit and fine, on both the actual third player behavior and the beliefs on corruption of the other two trust game agents. Fourth, we look at the impact of our treatments on a wide range of variables such as trust, the conditional distribution of trustworthiness (elicited with the strategy method), trustor's first order beliefs and strategic altruism (by measuring trustor's expected return on giving based on the expected reply of the trustee and behavior of the third player).

The well-known trade-off between lab experiments and standard econometric analysis passes through external consistency and capacity of isolating causal relationships. We accurately design our experiment to limit problems related to the first point. First and foremost, the design exactly reproduces the consequence of corruption under the form of bribery (i.e. a "tax" on the monetary payoffs of economic agents) subject to a risk of audit and penalty. Second, the standard investment game design is modified by introducing an effort task for the three players where the player with the worst performance becomes the third agent. This is done to foster the perception (beyond the information on the probability of audit and the penalty given by the experimenter) that the third player is withdrawing part of the other two players' payoffs without deserving it. Third, by modelling the probability of audit and the penalty, we clearly give the idea to all players that what the third player does is close to a bribery (and not just to legal taxation). Even though we are clearly aware that external consistency in a lab cannot be perfect, we believe that, for the reasons explained above, the design can reasonably capture some of the main elements involved in a fiduciary relationship in the presence of risk of extortion.

We use our experiment to test several null hypotheses on the three players' behavior under the standard assumption of purely self-regarding preferences being common knowledge. Our findings show that all the formulated null hypotheses are rejected. We observe a slightly higher level of trust in presence of corruption, matched by a significant effect of excess reciprocity from the trustee. Both the trustor and the trustee expect on average corruption acting as a tax, inelastic to changes in the risk of corruptor audit. Expectations are correct both for the inelasticity assumption, and for the actual value of the "corruption tax". We comment our findings explaining how they are consistent with an alternative view of human preferences where commendable giving triggers excess reciprocity, thereby justifying role and emphasis on heroes in all cultures and traditions around the world.

The rest of the paper is organized as follows. In the second section, we present the experiment design. In the third section our empirical findings are separately examined for trustors, trustees and third corrupting agents. In the fourth section, we conclude and resume the main findings and their implications for the literature

2. The experimental design.

2.1. Baseline game and additional treatments.

The baseline treatment, TC (trust + corruption), consists of a sequential game involving three players, A, B and C (Figure 1). Let q_A, q_B, and q_C denote the choices of the three players, respectively. At the beginning of the game, all of the three players are endowed with an initial, exogenous endowment, $E > 0$. In the first two stages of the game, A and B participate in a standard trust game. Specifically, in the first stage, A chooses how much of the endowment E to send to B, with $0 \leq q_A \leq E$. Whatever A sends to B is multiplied by a coefficient $\alpha > 1$, so that the amount effectively received by B at the end of the first stage is $\alpha q_A \geq 0$. In the second stage, B chooses how much to return to A of that amount, with $0 \leq q_B \leq \alpha q_A$. Thus, at the end of the second stage, the (temporary) payoffs of A and B from the trust game are given by $\pi_A^{trust} = E - q_A + q_B$, for A, and $\pi_B^{trust} = E + \alpha q_A - q_B$, for B.

Notice that, regardless of the choice made by B, the sum of A's and B's payoffs from the trust game is $\pi_A^{trust} + \pi_B^{trust} = 2E + (\alpha - 1)q_A$, thus uniquely determined by the size of the exogenous, initial endowment, E, and the choice of A. Let $S = (\alpha - 1)q_A$ denote the "surplus" generated in the trust game, namely the increase in the overall amount of resources at stake in the interaction. By the previous considerations, S is a function of the choice of the trustor, A.

In the third stage of the game, the corruptor, C, chooses how much of the surplus S to keep for herself, with $0 \leq q_C \leq (\alpha - 1)q_A$. Whatever C decides to keep reduces A's and B's payoffs from the trust game in proportion to the share of surplus, S, acquired by the subject in the trust game. More specifically, given q_C, the reductions in payoffs imputed to A and B are expressed by $\theta_A q_C$ and $\theta_B q_C$, respectively, where $\theta_A = max\left[\frac{\pi_A^{trust} - E}{S}; 0\right]$, $\theta_B = max\left[\frac{\pi_B^{trust} - E}{S}; 0\right]$, and $\theta_A + \theta_B = 1$.

C's choice completes the sequential game and final payoffs are $\pi_A^{TC} = \pi_A^{trust} - \theta_A q_C$, $\pi_B^{TC} = \pi_B^{trust} - \theta_B q_C$, and $\pi_C^{TC} = E + q_C$, for A, B and C, respectively.

We compare results from the baseline game TC with a modified version introducing the possibility for C of withdrawing part of the surplus produced by the other two players at the end of the game.

In the base treatment, TNC (Trust + No Corruption) the third agent, C cannot make any choice and simply receives feedbacks about the size of the surplus generated in the trust game. Thus, in TNC, the final payoffs of the three players are given by $\pi_A^{TNC} = \pi_A^{trust}$, $\pi_B^{TNC} = \pi_B^{trust}$, and $\pi_C^{TC} = E$ for A, B and C, respectively.

In the corruption treatment, we study the effects exerted by the opportunity given to the third player C of withdrawing part of the surplus generated by the other two players (with different audit probabilities). In particular, before making their choices in the TC_p (Trust + Corruption + Audit with probability p) treatment, all subjects are informed that, at the end of the game, C's choice will be audited with positive probability p. More specifically, we have three different versions of the corruption treatment with p respectively set at 0, 10 percent and 50 percent. The audit procedure influences C's final earnings only, while the expression of the A's and B's payoffs remain unchanged with respect to those in TC. In particular, if audited and found to have kept a positive amount of the surplus S, the payoff of the corruptor, C, is reduced by q_C plus a sanction that is proportional to the size of q_C. Specifically, the expected final payoff of C in TC_p is given by $\pi_C^{TC\_p} = (1-p)(E+q_C)+p(E-q_C-fq_C)$, where $f>0$ represents the flat fine rate that the corruptor pays on q_C if audited.

2.2. Procedures

Upon their arrival in the laboratory, subjects were randomly assigned to a computer terminal. At the beginning of the experiment, subjects were randomly and anonymously assigned to groups of three and were given a general description of the sequential game. In particular, subjects in TC_p were told that the interaction involved the following three phases (Figure 2):

Phase 1: the "slider task" competition. In the first phase, the three roles, A, B, and C, were assigned to group members depending on subjects' relative performance in the "slider task" real effort game (Gill and Prowse, 2012). The task consists of a single screen displaying a number of "sliders" and the layout of the screen was kept unchanged across subjects and sessions. All of the sliders on the screen were initially positioned at the left margin, which corresponds to the value of 0. By using the mouse, the subject could change an unlimited number of times the position of each slider at any integer location between 0 and 100 inclusive. She got a point whenever she manages to position a slider to the value of 50. At the end of the slider task, the score obtained by each subject is given by the number of sliders she centered to the value of 50 within an allotted time of 120 seconds. As the task proceeded, the screen displayed the subject's current points score and the amount of time remaining. At the end of the first phase, subjects were only informed about their final score and whether it ended up in the two best performances of the group. The two best performers in the group were randomly assigned to either role A or role B, while the subject with the lowest score was assigned to role C.

It might be argued that assigning roles on the basis of performances in a slider task might cause selection into roles. However, both the simplicity of the task and the fact that no specific skills (field of study, professional profile, etc.) were required to subjects to properly perform, reduces the extent of the potential bias from selection (our point is confirmed by empirical evidence on balancing properties discussed in section 3). On the contrary, two considerations justify this procedure. First, competing in the slider task could introduce a sense of entitlement to participate in the trust game, rather than being relegated to the mere role of C who either made no choice in the experiment (in TNC) or could only subtract resources from A and B (in TC and TC_p). Second, by making the difference between best and worst performers salient, the slider task enhanced the existing conflict of interest between participants in the trust game, A and B, and the corruptor, C.

Phase 2: the trust game. At the beginning of the second phase, each of the three subjects of the group were assigned to an exogenous endowment of ten tokens, namely $E = 10$. Then A and B participated in the trust game with the efficiency parameter set to $\alpha = 3$. The strategy method was used in order to elicit B choices. In particular, before being informed about A's decision, B chose how much to return to A for each possible choice that A could have made. B knew that, of all her 10 choices (one for each of the admissible values of q_A in $\{1,2, \dots ,10\}$), only the choice corresponding to the amount effectively sent by A would have been used to determine payoffs. The decision of using the strategy method for eliciting B's choice allows us to collect detailed information about B’s conditional behavior and to assess whether her attitude to reciprocate is affected by the presence of the corruptor and the existence of formal rules of auditing and prosecution.

Phase 3: the choice of C and the audit procedure. The choice of C is also measured with the strategy method. Namely, the player chose how much of the surplus S to keep for any possible choice of A. Again, C knew that, of all her 10 choices (one for each of the admissible values of S in $\{2,4, \dots ,20\}$), only that corresponding to the amount effectively sent by A would have been used to determine payoffs. The audit procedure was administered by the PC. In particular, subject C was told that the computer would have randomly selected one of 100 tickets, numbered from 1 to 100. If the number of the ticket was smaller than p, then the choice of C was audited. We used two values of p, either 10 or 50, thus generating two treatments: TC_50, in which the probability of auditing was fifty percent, and TC_10 in which the probability was ten percent. In both treatments, we set the penalty rate $f = 0.5$, implying that, in case of auditing, C was convicted to pay a fine of one token for every two subtracted from the surplus, S. The value of f was chosen in order to avoid bankruptcy of C.

All the relevant rules to determine payoffs, as well as information about how C's choice affected A's and B's payoffs, were provided at the beginning of the experiment and included in the general description of the game. Specific instructions for each of the three phases were distributed at the beginning of each phase and read aloud. Before starting with each phase, subjects answered to a number of control questions to assure about their understanding of the experimental rules and, in case of necessity, they were assisted by researchers in order to ensure full understanding of such rules.

Finally, we elicited belief measures about B's and C's choices by using an incentive compatible mechanism. In order to minimize the risk of hedging between choices and beliefs, subjects were informed about the belief elicitation procedure after phase 3 and before receiving feedbacks on subjects' choices and payoffs. Again, instructions were distributed and read aloud.

A's first order and B's second order beliefs in the trust game. We collected A's guesses on the amount returned by B and B's guesses on the amount A's expected her to return in the trust game. Since B made her choice in strategy method, A and B were asked to state beliefs for each of the 10 potential choices that B could make. A and B were also told that, at the end of the experiment, one of the ten guesses would have randomly picked by the computer. In case the selected conjecture turned to be correct, they received three tokens in addition to the payoff of the three phases.

A's and B's first order beliefs about C's choice. We collected A's and B's guesses on the share of surplus generated in the trust game that was kept by C in the third phase of the experiment. Again, since C made her choice in strategy method, A and B were asked to state beliefs for each of the 10 potential choices that C could make. As before, at the end of the experiment, the computer randomly selected one conjecture and A and B were paid 3 tokens according to its correctness.

The only difference between TC_p (p=0) and TC_p (p>0) concerned the fact that, in the former, we removed the audit procedure and, therefore, C did not face the risk of audit. Instead, the difference between TC_p and TNC was that, in the latter treatment, C did not make any choice and only observed the surplus generated in the trust game. In all treatments, the language used in the instructions was kept as neutral as possible and did not refer to sensitive words, such as “trust”, “corruption”, etc. (an English version of the experimental instructions used in TC_50 are included in Appendix A).

Subjects were informed that, during the experiment, payoffs and choices were expressed in tokens, rounded to the closest integer, if necessary. At the end of the session, the number of tokens accumulated during the experiment were converted at an exchange rate of 1 euro for 2 tokens and monetary earnings were paid in cash privately. On average, experiment participants earned about 9.59 euros (including 3 euros for showing up) for sessions lasting about 45 minutes, including the time for instructions and payments. Before leaving the laboratory, subjects completed a short questionnaire containing questions on their socio-demographics and their perception of the experimental task. We ran 3 sessions per treatment, each involving 15 subjects, for a total of 180 participants. The experiment took place in June 2017 in the Behavioral Economics Research Group (BERG) laboratory of University of Cagliari. Participants were randomly recruited from the BERG subject pool, which consists of approximately 1000 students from a wide range of disciplines. The experiment was computerized using the Z-Tree software (Fischbacher, 2007).

2.3. Hypothesis testing

As is well known, with common knowledge on purely self-regarding preferences, the Nash equilibrium of a trust investment game is the “no investment”, “no return” situation, that is, the pair of strategies where the trustor gives 0 and the trustee returns 0. The situation becomes different if the assumption of common knowledge of purely self-regarding preferences is relaxed. In such case, if the trustor believes that the trustee is other-regarding and that trust pays (i.e. the trustee will return more than what initially sent), she may find it optimal to give a strictly positive amount of resources.

Conversely, the choice of a purely self-regarding trustee, does not change conditionally to the assumption of the counterpart's (trustor's) preferences. Her optimal conditional giving is null, even though she expects nonzero giving from the trustor.

Optimal strategies of the trustor and the trustee are as well unaffected by the introduction of the third "corrupting" player in our modified trust investment game in three cases (trustor giving with common knowledge on purely self-regarding preferences, and trustee giving with or without common knowledge on purely self-regarding preferences). Under the fourth case (trustor gives a positive amount of resources under the expectation of other regarding preferences from the trustee), the choice depends on the effect of the third player withdrawal (acting as a tax on the surplus created by the trustor in the game). There may be cases where the strategic altruism of the purely self-regarding trustor pays without the tax, but it is expected not to pay anymore with the tax since the gross trustor 's payoff (the trustors' payoff net of tax) is expected to be higher (lower) when giving than when not giving. In these cases, the introduction of the third agent in the game modifies (reduces) trustor giving.

Last, it is trivial to observe that, in case of purely self-regarding preferences, the third agent withdraws a 100 percent tax on the surplus produced by the other two players. Exactly as we expect zero giving from a purely self-regarding player in dictator games.

These considerations lead us to formulate the following three standard null hypotheses for our research:

H1: $(q_A = 0 \; in \; TNC \; and \; TC\_p)$. *When the trustor is purely self-interested, and this is common knowledge, the possibility of withdrawal of part of the surplus from a third "corrupting" agent in the trust game has no effect on trustor's behavior. Trustor giving is zero both with and without corruption*

A variant of this hypothesis is that, when trustors believe that trustee may give non-zero (purely self-regarding preferences are not common knowledge), but believe that trust does not pay they will still choose zero giving based on their purely self-regarding preferences.

H2: $(q_B = 0\ in\ TNC\ and\ TC\_p)$. *In a setting with purely self-regarding trustees the possibility of withdrawal of part of the surplus from a third "corrupting" agent in a trust investment game has no effect on trustee's giving (with or without common knowledge on purely self-regarding preferences). Trustee's conditional giving is zero, both with and without corruption.*

H3: in a setting with purely self-regarding trustors, where trustors expect that trustees are non-purely self-regarding, the possibility of withdrawal of part of the surplus from third "corrupting" agent in a trust investment game has effect on trustor's behavior (from zero to nonzero giving) if the trustor expects that his/her giving is expected to pay (not to pay) without (with) the tax (where for "pay" we mean trigger a return higher than the amount sent)

Note that the previous hypothesis is based on the idea that trustors expect that trust pays in treatments without corruption, while it does not in treatments with corruption. To verify this conjecture, we need to calculate trustor's expected return from giving as we do in section 3.1.3.

H4: If the third agent has purely self-regarding preferences she will apply a 100 percent tax on the surplus created by the other two players in the corruption treatments without penalty

The expected payoff for the third agent in case of withdrawal is $p_wS(1-p)-(p_wS/2)p$ where p_w is the share of the surplus withdrawn and p the probability of being detected. The third agent, even if risk averse, will charge the highest (100) tax in corruption treatment without penalty (since she does not run any risk of being prosecuted) and in corruption treatments with penalty if she is not risk averse (since a higher tax will raise her expected payoffs). If she is risk averse and if her level of risk aversion is high, she may prefer less than 100 percent tax in treatments with positive probability of audit since the decision to withdraw is equivalent to running a bet of increasing risk as far as the withdrawal rate is higher.

3. Empirical findings

We start our empirical analysis by testing whether balancing between corruption and non-corruption treatments is guaranteed (see Table 1). We find a one year difference between participants playing treatments with/without corruption (not significant at 95 percent), while the gender balance is almost the same (44 percent against 40 percent in corruption versus no-corruption treatments). The share of players doing voluntary activity is 44 against 51 percent in no-corruption versus corruption treatments (the difference is again not statistically significant). University years are on average three in both groups, even though students range from first year undergraduate to second year graduate. The share of those knowing game theory is almost the same in corruption and no-corruption treatments, and approximately one third. We as well do not find significant differences in gender, age, self-assessed risk aversion, voluntary status and share of players knowing game theory in two-by-two role comparisons (ie., trustor versus trustee, trustor versus third corrupting agent, etc.). This finding confirms that selection into roles is not a problem in our experiment. To conclude with, there are no relevant and significant differences in socio-demographic characteristics among players in corruption and no-corruption treatments and in two-by-two role comparisons. Consequently, we can conclude that balancing properties are met in our between design experimental setting.

3.1. Trustor behavior

A first inspection of the behavior of our 60 trustors under the different treatments does not reveal, apparently, any significant effect of corruption on their behavior. From a descriptive point of view, trustors surprisingly give slightly more in corruption treatments and, specifically, 5.6 ECUs in the treatment with zero probability of auditing of the third corrupting agent, 4.53 in the treatment with 10 percent probability of auditing, and 4.66 in the treatment with 50 percent probability of auditing. Average trustor giving in the no corruption treatment is 3.87 ECUs (Table 2). The difference between corruption and no-corruption treatments is however not significant under both parametric and non-parametric tests (two-sided Mann-Whitney rank-sum test, z 1.31, p-value 0.19). Note that this finding does not necessarily imply rejection of the hypothesis of trustor's purely self-regarding preferences under Ho(1), since evidence on trustor giving *per se* is uninformative about the common knowledge assumption, that is, we do not know whether the trustor expects that trust pays or not (we will verify this point in sections 3.1.1 and 3.1.2). In the first case, she can still have purely self-regarding preferences and opt for nonzero giving.

With a deeper inspection of our data, using also information from post-experiment questionnaire, we find that 40 trustors declare they have never studied game theory, while 20 of them declare they did. If we look at the effect of game theory knowledge on average trustor giving in games with corruption, we find that trustors who declare not to know game theory give significantly more (5.52 ECUs, against 3.34 of those who know game theory) (Table 2). The non-parametric test on the difference in giving between the two groups of trustors created when using game theory knowledge as a discriminating factor, rejects the null (z=2.34, p-value 0.019 in the two-sided Mann-Whitney rank-sum test).

The problem here is that knowledge of game theory is not a randomized variable in our experiment. What we observe may therefore well be endogenous and due to a sorting and matching effect (i.e. individuals with stronger pro-social attitudes decide not to study game theory). This implies that, based on our evidence, we cannot proof that studying game theory reduces *per se* trustor giving. Nonetheless, the observed significant correlation is of great interest suggesting that heterogeneity in this specific population characteristic can produce relevant effects. We ask in another part of our post-experimental questionnaire what was the strategy followed by players (i.e., maximizing their own payoffs, maximizing payoffs of the team, etc.) The share of individuals declaring that they acted maximizing their own payoffs was 50 percent among those who studied game theory, while just 18 percent among those who did not study it. The same two shares related to those declaring that they acted maximizing the team payoff were 13 percent and 28 percent. We can therefore associate knowledge of game theory to the prevalence of a purely self-regarding approach and ignorance of game theory to a different, more other-regarding (or we-thinking) attitude.

3.1.1. Trustors' expectations on third agent withdrawal

In order to check whether the different behavior of trustors, according to knowledge of game theory, was affected by pure or strategic altruism, we look at trustors' expectations about the behavior of the third corrupting agent, and at their beliefs on the trustee's strategy.

Specifically, on the expectation about the behavior of the third corrupting agent, we estimate the following specification

$$E[CorruptionTax]_{ij} = \alpha_0 + \alpha_1 Surplus_{ij} + \alpha_2 Surplus_{ij} * DTC10 + \alpha_3 Surplus_{ij} * DTC50 + \alpha_4 Surplus_{ij} * D1GAME + \alpha_5 Surplus_{ij} * DTC10 * D1GAME + \alpha_6 Surplus_{ij} * DTC50 * D1GAME + \eta_i + \varepsilon_{ij}$$

(1)

where the dependent variable $E[CorruptionTax]_{ij}$ is the *i*-th trustor expectation of the third agent's withdrawal conditional to the *j*-th trust game surplus ,. On the right hand side the trust game surplus

(*Surplus*) is equal to twice the trustor giving (see section 2), *DTC10* (*DTC50*) are the two (0/1) dummies for treatments where the probability of audit is 10 percent (50 percent), respectively, and *D1GAME* is a (0/1) dummy taking value one for trustors who declare to know game theory in the post experiment survey. The specification is estimated with fixed effects (η_i)[1] and is obviously limited to observations collected in corruption treatments. Note as well that the few socio-demographic characteristics that could matter (age, gender) are time invariant, are therefore absorbed by fixed effects and not specifically introduced in the specification.

Estimate findings show an average trustors' expectation of a "corruption tax" (third player's withdrawal rate) of around 62 percent, inelastic to differences in the probability of audit and fine of the corruptor (see Table 3, column 1). However, if we introduce dummies for the expectation of trustors who do know game theory, we find that the latter expect a higher tax (almost 100 percent) in the treatment with corruption and zero probability of audit, while a significantly lower tax in corruption games with probability of audit (see Table 3, column 2). As we will see in section 3.3, the guess of trustors who know game theory reveals to be too pessimistic and that of trustors who do not know game theory closer to the actual third agent's withdrawal. This is because the third agent will "levy" a 70 percent tax (not significantly different from 62 percent however, given the variability of forecasts, when using 95 percent confidence intervals), inelastic to the probability of audit. Fixed effects are, as expected, strongly significant, and their inclusion allows us to estimate properly the within effect of changes in the generated surplus on the expected third agent's withdrawal.

3.1.2. Trustors' expectations on trustee's giving

[1] Since group members never interacted with other participants in the session, we run these regressions (and those that follow) without controlling for potential dependency across subjects in different groups. We nevertheless repeated regressions with either clustered standard errors to control for potential sources of dependency within session, or robust bootstrapped standard errors to control for non-normality and heteroscedasticity of residuals. Results remain almost unchanged and are available upon request.

In order to investigate the determinants of the trustors' expectations on trustee's giving we estimate the following specification

$$E[Payback]_{ij} = \alpha_0 + \alpha_1 TrustorGive_{ij} + \alpha_2 TrustorGive_{ij} * DTC00 + \alpha_3 TrustorGive_{ij} * DTC10 + \alpha_4 TrustorGive_{ij} * DTC50 + \alpha_5 TrustorGive_{ij} * DTC00 * D1GAME + \alpha_6 TrustorGive_{ij} * DTC10 * D1GAME + \alpha_7 TrustorGive_{ij} * DTC50 * D1GAME + \eta_i + \varepsilon_{it}$$

(2)

where the dependent variable *E[Payback]*$_{ij}$ is the i-th trustor expectation on the trustee's payback, conditional to his own j-th amount of giving. With this variable, we therefore measure trustors' beliefs over the trustees' strategy method that we record in our experiment. *TrustorGive* is a given amount of trustor giving assumed by the trustee in the strategy method and the other right-hand-side variables are defined as in (1).

Findings from (2) show that, on average, trustors predict a 1.23 payback ratio falling to around 1 in games with corruption and not highest (50 percent) probability of audit (Table 4, column 1). This means that trustors expect on average that trust will pay in terms of reciprocity (net of the intervention of the third corrupting agent) only without corruption, or with corruption and high probability of audit. They as well expect that trust will not pay in games with corruption and zero or mild (10 percent) probability of audit. If we however decompose the effect in separate estimates for trustors who know and do not know game theory, we find that the payback ratio without corruption is much lower (around 1.06) for the former, falling substantially (by around 0.3) in games with corruption and not highest probability of audit (Table 4, column 2). In synthesis, trustors who do know game theory expect trust to pay almost nothing in all treatments, and to lose in those with corruption and mild probability of audit. They nonetheless choose nonzero giving in these last treatments and these combined findings therefore lead to the rejection of Ho(1), implying that the above mentioned trustors cannot have purely self-regarding preferences.

Differently, trustors who do not know game theory have rosier expectations on trustees' strategies, predicting an average payback ratio of 1.43 that falls by around .37 in games with corruption, and not highest probability of audit (Table 4, column 3). In synthesis, for them trust pays in terms of reciprocity (and weakly so even in games with corruption and not highest probability of audit). In Table 4, column 4 we simply re-estimate the model for the overall sample of trustors introducing slope dummies in order to check whether the observed differences between trustors who know and do not know game theory are significant or not. Findings from this estimate show that the above described differences in expectations between the two groups are significant.

3.1.3. Trustors' expectations on returns of their strategies

In order to make clearer the link between trustor giving and her expectations on trustee's and third agent's behavior (described in the previous two sections) we calculate the expected return on giving based on the above mentioned two expectations (separately looking at agents knowing or not game theory).

The expected return is measured as

$$E[R]_i|_{Tk}=[q_A+S-E[q_B]-\vartheta_A E[q_C]]/\ q_A$$

where Tk is the specific treatment and k = NTC, TC00, TC10, TC50 and the other variables are described as in section 2.1.

Our findings show that players who know game theory expect that trust pays (with a 6 percent return) only in treatments without corruption, but they surprisingly choose nonzero giving also in treatments with corruption where they expect that trust does not pay (Table 5). This implies rejection of H1 since their giving is significantly different from zero. More specifically, average giving of trustors who know game theory in treatments with corruption is 3.34 (t-stat 5.08, p-value 0.000, non-parametric test z 2.97, p-value 0.003), with n=12 and only 2 players out of 12 giving zero. Note that our observations are very limited when we look at the game theory knowledge/treatment type subgroups, so that some of the findings provided in Table 4 are purely descriptive. Players who do not know game theory expect that trust always pays (from a maximum of a 43 percent return in the treatment without corruption, to a minimum of a 2.9 percent return, in the treatment with corruption and 10 percent probability of audit). Consistently, their levels of giving are generally higher than those of players who know game theory but, again, surprisingly higher in treatments with corruption than in those without it.[2]

3.1.4. Discussion on trustors' behavior

Evidence provided in sections 3.1.1-3.1.3 shows that trustors do not behave following the Nash prediction of zero giving, neither in the standard trust game without corruption, nor in the game with corruption. This is only partially justified by their average belief of higher than unit trustees' payback ratios where their non-zero giving may be explained by strategic altruism. This is the case of trustors who do not know game theory and expect on average higher than unit payback ratio (and not a 100 percent corruption tax) both in corruption and no corruption treatments. However, we also observe that players who know game theory expect lower than unit payback ratios from trustors and an almost 100 percent corruption tax, and still choose higher than zero giving. Their average giving in treatments with corruption is indeed 3.34 (and close to it even when we exclude the treatment with corruption and highest probability of audit). This result leads to the rejection of H3 since, when expected returns switch from negative to positive (from corruption to no-corruption treatments) we do not find a significant increase in giving. Quite to the contrary, average giving moves from 3.34 to

[2] The extremely limited number of observations for each intersection explains the lack of consistency of behavior when moving treatments of lower to higher probabilities of audit (ie., there is only one player who knows game theory and participates to a treatment with corruption and highest probability of prosecution).

2.12 (Table 5). The finding cannot therefore be explained by strategic altruism, while it can be explained by other-regarding components (only 50 percent of players who know game theory declare ex post they acted in the game maximizing their own payoffs and not taking into account as well the interest of other players). To sum up, the prediction of a 62 percent flat tax of the corrupting agent in treatments with corruption does not reduce, but actually slightly increases giving in players who do know game theory but also in those who do not know it (from 5.52 to 4.25). This is another “anomaly” that will be matched by a similar excess giving from trustees in treatment with corruption (see the next section).

3.2.Trustees’ behavior

In order to analyze trustees’ behavior we start by estimating their expected withdrawal from the third agent (corruption tax) with the following specification

$$E[CorruptionTax]_{ij} = \alpha_0 + \alpha_1 Surplus_{ij} + \alpha_2 Surplus_{ij} * DTC10 + \alpha_3 Surplus_{ij} * DTC50 + \alpha_4 Surplus_{ij} * D2GAME + \alpha_5 Surplus_{ij} * DTC10 * D2GAME + \alpha_6 Surplus_{ij} * DTC50 * D2GAME + \eta_i + \varepsilon_{ij} \quad (3)$$

where the dependent variable now measures the the *i*-th trustee’s expectation on the third agent’s withdrawal conditional to the j-th surplus expected by the third agent in the strategy method, *D2GAME* is a (0/1) dummy taking value one if the trustee declares ex post to know game theory and the other variables are defined as in (1).

Findings from (3) show that, on average, trustees expect a corruption tax of 64 percent (Table 6, column 1). When we estimate the specification separately for trustees who know and do not know game theory, we find that the former are slightly more pessimistic (67 percent against 61 percent) (Table 6, columns 2-3).

The analysis of trustees' choices in the strategy methods with the specification below allows us to compare their actual behavior with what predicted by trustors (equation (2) in section 3.1.2). The estimated specification is

$$Payback_{ij} = \alpha_0 + \alpha_1 TrustorGive_{ij} + \alpha_2 TrustorGive_{ij} * DTC00 + \alpha_3 TrustorGive_{ij} * DTC10 + \alpha_4 TrustorGive_{ij} * DTC50 + \alpha_5 TrustorGive_{ij} * D2NOGAME + \alpha_6 TrustorGive_{ij} * DTC00 * D2NOGAME + \alpha_7 TrustorGive_{ij} * DTC10 * D2NOGAME + \alpha_8 TrustorGive_{ij} * DTC50 * D2NOGAME + \eta_i + \varepsilon_{it}$$

(4)

The dependent variable (*Payback*) is now the actual payback of the *i*-th trustee conditional to the *j*-th trustor's contribution, *DNOGAME* is a (0/1) dummy for trustees who declare ex post they do not know game theory and the other variables are defined as above. In alternative to (4), we use in a slightly modified specification a unique dummy for corruption treatments (*DCORRUPTION*) and *D2GAME* as a (0/1) dummy taking value one if the trustee does know game theory, in order to test the effect of the two variables on trustees' strategy (Table 7, column 5)

Findings from (4) show that trustees' payback ratio (payback/trustor giving) in the aggregate sample (when considering all players in corruption and no-corruption treatments) is 0.67, but it is higher in games with corruption (actually 0.20 higher in games with zero probability of capture, while 0.40 in games with 50 percent probability of capture) (Table 7, column 1). These findings clearly show that trust does not pay but they also lead to the rejection of H2 in two directions: i) trustees do not choose zero giving in treatments with/without corruption; ii) giving in treatments with corruption is not the same as giving in treatment without corruption, and is actually higher. Note that this excess reciprocity effect produced by the introduction of corruption in our treatment is a controlled experimental result, valid on the overall sample for all trustees (irrespective of declared characteristics such as knowing or not knowing game theory and/or declared strategies followed in the game). An immediate conclusion drawn when comparing this finding with those of the previous section is that trustors on average overestimate trustees' giving.[3]

If we analyze trustees' giving separately, for trustees that know game theory and those who do not (37 percent of trustees[4]), we find that the effect is driven by the former who raise significantly their payback ratio in corruption treatments (Table 7, columns 2 and 4). On average, the payback ratio falls significantly from the above mentioned overall sample average, by .46 for trustees who know game theory in corruption treatments, while it raises by .36 percent by those who do not know it (Table 7, column 5). As a final remark on this point, consider that most of the excess reciprocity effect occurs in corruption treatments with the highest probability of audit. Hence, effectiveness of prosecution plays an important role in determining the resilience of trustworthiness in difficult economic environments.

Note that this finding is only partially driven by differences in expectations about the third agent's withdrawal. This is because, when we test whether knowing game theory produces a difference in such expectations among trustees, we find that the null is not rejected (Table 6, column 4). However, when we estimate the regression separately for trustees who know and do not know game theory, we find a 6 percent difference in the predicted corruption tax, with trustees who do not know game theory being more optimistic (Table 6, columns 2 and 3).

To sum up, the paradoxical experimental finding of our trust game with corruption on the trustee side is that trustees give significantly more in presence of corruption treatments (excess reciprocity effect of corruption), even though they expect that the corruptor will charge a flat tax of 62 percent on the

[3] Beyond this general finding however, we know that trustors who know game theory actually underestimate it in treatments with corruption, while trustors who do not know game theory overestimate them in treatments without corruption (see Table 4).

[4] Decision to maximize one's own payoff is declared ex post by 45 percent of trustees knowing game theory against 32 percent of those not knowing it.

surplus generated by the trust game. We also find that our result is driven by trustees that do not know game theory, even though the latter have expectations on the corruption tax not significantly different from those of trustees that do know game theory, and still expect a strong corruption tax with an expectation not significantly different from the reality. A likely interpretation is that trustors who do not know game theory follow a reciprocity rule that includes a premium for trustor giving in more difficult “environmental” conditions such as those of the treatments with corruption.

The excess payback of trustees matches the excess giving of trustors, thereby configuring a sort of gift exchange in treatments with corruption that has the effect of raising trust and trustworthiness in a difficult economic environment where players know that part of the surplus they generate can be lost. Note that such behavior cannot be motivated by scarce knowledge of the game since players correctly expect that part of their surplus will be taken by the third corrupting agent.

3.3. Interpretation of trustees’ excess reciprocity

The experimental findings of the higher trustee’s giving in treatments with corruption are in strong contradiction with the standard hypothesis (H2) formulated in section 2.1. They may be however consistent with a different definition of players’ preferences. Consider the importance of heroes (in the non-religious culture) and of saints (in the religious culture). Why so much emphasis has been placed in the past on them? One of the reasons was that the celebration of virtues of heroes and saints was meant to stimulate virtues of the common people in order to make easier the achievement of some societal or institutional goals. The translation of this idea in terms of preferences implies the hypothesis of an excess component of reciprocity that could be activated in case of observation of an action of another human being that was particularly commendable. Our modified trust game treatment can be used to test whether human populations possess this excess reciprocity characteristics. This is because the difference with respect to a standard trust game is that trustor’s giving can be considered relatively more praiseworthy by trustees that know that, in case of non-zero third agent withdrawal, part of that giving will get lost for them. If the specific component of reciprocity preferences awarding praiseworthy behavior or emulating it exists and is strong, it could produce a significantly higher trustee's giving in treatments with corruption.[5] And this is what we observe. Note as well that the idea that a particularly praiseworthy behavior of trustors may trigger excess generosity from the trustee is consistent with the same findings of the first experimental trust game results in Berg et al. (1995)[6] and with behavioral principles such as guilt aversion (Dufwenberg, 2002; Battigalli and Dufwenberg, 2007) and trust responsiveness (Guerra and Zizzo, 2004; Bacharach et al., 2007; Pelligra, 2005 and 2010).

3.4. Third agent's behavior

Findings discussed in the previous sections show that both trustors and trustees believe that the presence of the third "corrupting" agent will significantly change their payoffs. They in fact expect, on average, that the corruptor will take around 63 percent of the surplus generated by trustor's giving, (Table 8). We as well know that expectations of trustors who do not know game theory are rosiest (Table 3, column 2). The 63 percent average expectation on the third agent's withdrawal is significantly different from zero in the fixed effect estimate of the expected conditional corruption tax including both trustors and trustees (95 percent confidence interval ranging from 57 percent to 69 percent). Note that in this estimate the hypotheses of changes in the tax according to differences in the risk of audit are rejected, hence the corruption tax is expected to be inelastic to the probability of audit. We also test whether trustees and trustors beliefs differ and find that this is not the case.

[5] Of course, in the hero-common people reaction scheme the excess reciprocity is not directed toward the hero, but is addressed to its same cause. This does not change the essence of our reasoning.

[6] The authors find that trust does not pay on average in their 32 observations (average amount sent 5.16, average payback 4.66). However, for trustors sending the entire amount (10 ECUs) trust did pay (average payback 10.2).

We now compare these expectations with the actual third agent's behavior. We estimate the latter with the following specification

$$CorruptionTax_{ij} = \alpha_0 + \alpha_1 Surplus_{ij} + \alpha_2 Surplus_{ij} * DTC10 + \alpha_3 Surplus_{ij} * DTC50 + \alpha_4 Surplus_{ij} * D3GAME + \alpha_5 Surplus_{ij} * DTC10 * D3GAME + \alpha_6 Surplus_{\square j} * DTC50 * D3GAME + \eta_i + \varepsilon_{ij}$$

(5)

where the difference with (1) is that the dependent variable is now the actual withdrawal of the i-th third agent, conditional to the j-th surplus generated by the other two players and D3GAME is a (0/1) dummy taking value one if the third agent does know game theory.

Findings from (5) show that the guess of trustors and trustees is correct with regard to the inelasticity to the probability of capture, but it slightly underestimates the actual rule followed by the third agent, that actually uses a 71 percent tax (Table 9, column 1). The intercept coefficient in Table 8, column 1 is .708, with 95 percent confidence intervals .657 and .758, hence significantly different from the 100 percent withdrawal. The same coefficient in Table 8 column 2 is .638, with 95 percent confidence intervals .5665 and .710. In an estimate without regressors and just fixed effects and intercept we get an average tax of .707, with 95 percent confidence intervals .677 and .736, leading to the rejection of the null of 100 percent withdrawal. The same numbers when we limit the estimate to treatments with corruption and zero probability of audit are .708 with 95 percent confidence intervals .666 and .750, leading to rejection of the null also when third agents do not run the risk of being prosecuted. This last finding implies rejection of our fourth null hypothesis (H4) in section 2.3

It is as well remarkable that the 95 percent confidence intervals of the actual third agent behavior (62 / 79 percent) and the forecasted third agent behavior by all trustors and trustees (57 / 69 percent) overlap, thereby documenting a not statistically significant prediction error of trustors and trustees when considered in aggregate. Note as well that, as remarked above, we however know that, behind this aggregate result, trustors who do not know game theory are closer to the actual value than those who know it and overestimate the tax.

A final test is whether third agents knowing game theory behave differently than those who do not. We find that this is the case. The 71 percent corruption tax splits into a 64 percent tax levied by corruptors who do not know game theory, and a 77 percent tax levied by the complementary sample (Table 9, column 2). This is again a significant difference in our experiment driven by knowledge of game theory, with the latter bringing players to behavior closer to that of the purely self-regarding paradigm.

4. Discussion and conclusions

Our experimental findings on trust in a difficult (corrupt) economic environment reject all standard hypotheses based on the assumption of common knowledge and purely self-regarding preferences. We find ten main interesting results. Specifically: the different corruption treatments do not significantly change aggregate trustors' behavior, that is significantly different from zero in all cases (fact one). However, trustors without previous knowledge of game theory give significantly more in treatments with corruption with respect to trustors who know it (fact two). The former expect that trust pays (even though less in treatments with corruption), and therefore part of their behavior is driven by strategic altruism (fact three). Trustors who know game theory expect that trust does not pay in treatments with corruption, but they nonetheless give significantly more than zero (fact four). When looking at actual trustees' behavior we find that trust does not pay (since the payback ratio applied by trustees is around 62 percent) (fact five), consistently with what expected by trustors who know game theory. The introduction of corruption in the standard trust game experiment however generates significant excess reciprocity from trustees (fact six). The excess reciprocity result is driven by trustees who do not know game theory (almost 2/3 of the sample), but still expect (not differently from trustees who know game theory) that the third agent will take 62 percent of the surplus (fact seven). The third agent corruption tax is inelastic to the probability of audit (and correctly expected to be so by trustors and trustees, with the exception of trustors who know game theory) (fact eight). There is a slight (but not statistically significant) underestimation of the corruption tax by both trustors and trustees (fact nine), whose prediction on the behavior of the third agent is substantially correct. Third agents who know game theory charge a significantly higher tax on the other two players (fact ten).

These findings reject the four standard hypotheses on players' behavior under purely self-regarding preferences that have been discussed in section 2.3, while they are consistent with an alternative configuration of preferences that we argue as justifying the role of heroes in culture and past traditions. Such alternative configuration implies that individuals have an excess reciprocity component that can be triggered when they observe counterparts giving that can be considered particularly praiseworthy. This is the case in our treatments with corruption where trustor giving is more commendable given the risk of withdrawals of part of the surplus by the third "corrupting" agent.

We believe that, from these ten facts, two further fundamental lessons can be learned, opening way for future research.

First, at least three of our main findings provide evidence of trust and trustworthiness resilience in difficult economic environment (i.e. in presence of experimental treatments reproducing the main economic effects of corruption). This occurs because: i) trustors who do not know game theory give more in corruption treatments, partly because they expect more than unit payback ratios from trustees, even though lower than unit payback ratios in treatments without corruption; ii) trustors who know game theory (and expect a less than unit payback ratio from trustees and a higher corruption tax), choose nonetheless nonzero giving in corruption treatments; iii) the corruption treatment produces excess reciprocity (driven by trustees who do not know game theory). Facts i) and iii) outline a sort of “gift exchange” phenomenon, limited to players who do not know game theory. This gift exchange mechanism has the power of raising trust and trustworthiness under difficult economic environments. It is important to remark that such resilience is not produced by misunderstanding of the corruption added feature of the game, because expectations on the third corrupting agent’s behavior from experiment participants who do not know game theory are not statistically incorrect. It is also important to consider that a relevant part of this resilience is produced in corruption games with the highest probability of audit. Hence, effectiveness of prosecution accounts for an important part (even though not all) of trust and trustworthiness resilience in difficult economic environments.

Second, as it is clear from the above described findings, knowledge/ignorance of game theory matters in discriminating between purely self-regarding and other-regarding strategies, and in players’ expectations and actual behavior. Specifically, players who know game theory exhibit behavior and beliefs closer the purely self-regarding paradigm. As already discussed above, this is not an experimentally controlled factor in our research. This means that it is not possible to verify whether it is game theory knowledge, *per sé*, that produces the effect, or a sorting mechanism leading individuals closer to the purely self-regarding paradigm to follow studies including game theory in their curricula[7]. In spite of it, this specific finding tells us that is of foremost importance to take into account that populations are highly heterogeneous in educational background and preferences when modelling, investigating and predicting economic agents’ behavior. In this sense, the incorrect and too pessimistic beliefs on the third agents’ corruption tax by trustors who do know game theory may have been driven by the erroneous expectation that all third agents know game theory and behave following purely self-regarding preferences.

Our findings produce as well interesting inferences on the aggregate dynamics of corruption and growth if we regard the trust investment game as the microeconomic core of the process of creation of economic value. In our experiment corruption treatments yield higher gross output (by considering

[7] Previous studies, however, tend to support the former explanation (Bauman and Rose, 2011)

it as the sum of the traditional aggregate trust game “output”, including the part of the surplus taken by the third corrupting agent), but lower net “output” (the observed sum of payoffs of trustors and trustees after the corruption tax) vis-à.vis the output of no corruption treatments. This is consistent with an observed negative effect between corruption and growth (under the reasonable assumption that the corruption tax goes in the informal economy), even though the gross effect, when adding the corruptor’s take in the informal sector, may become surprisingly positive.

To conclude with, our experiment suggests that a combination of effectiveness in prosecution, other-regarding preferences and gift exchange mechanisms, where the commendable trustor giving in corruption treatments triggers excess reciprocity, may produce even in economic environments plagued by corruption and crime unexpectedly high levels of trust and growth, even though part of them are not measured by official statistics. At the core of this unexpected phenomenon we find “commendable” giving triggering excess reciprocity.

Figures and Tables

Figure 1. The Trust Game with Corruption

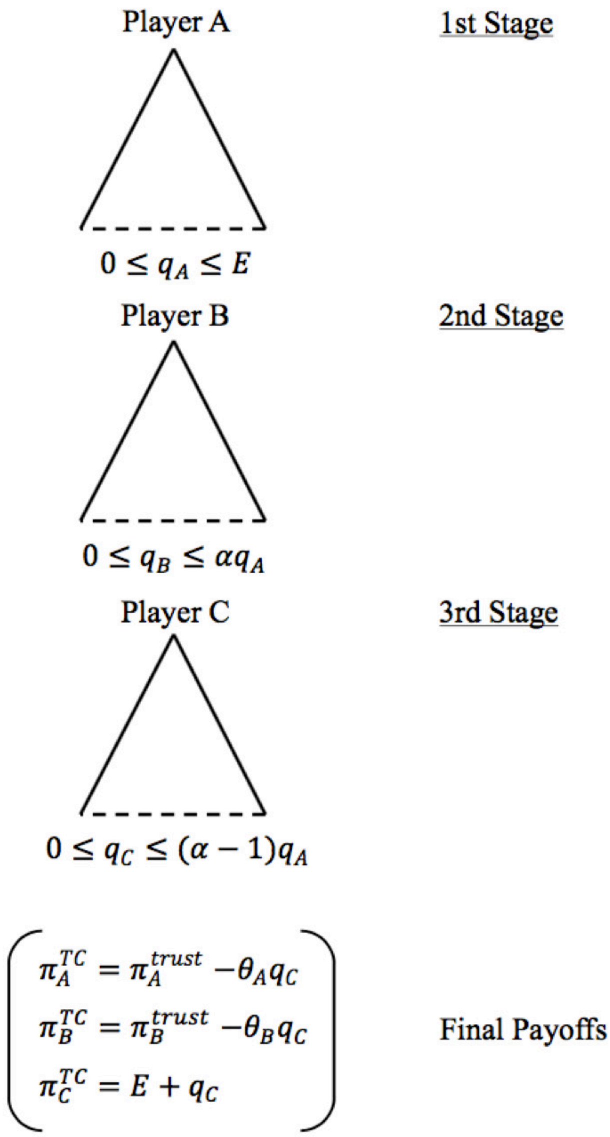


Figure 2. Timeline of the experiment

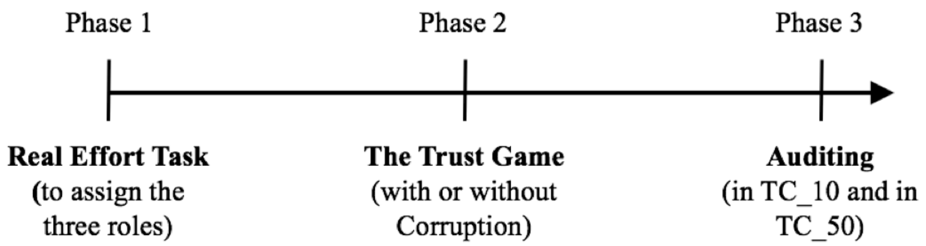


Table 1. Balancing properties (between corruption/non corruption treatments and players' roles)

	Corruption vs non-corruption treatments	Trustor vs trustee	Trustor vs third corrupting agent	Trustee vs third corrupting agent
Male	0.520 (0.603)	0.55 (0.585)	-1.68 (0.093)	-2.21 (0.027)
Age	2.23 (0.024)	-0.14 (0.887)	1.07 (0.204)	1.27 (0.203)
Voluntary status	0.53 (0.598)	-2.131 (0.033)	z =0.029 (0.977)	1.812 (0.070)
Years of education	-0.005 (0.998)	-0.961 (0.343)	-0.847 (0.402)	0.092 (0.932)
Risk aversion*	-1.624 (0.104)	-0.073 (0.941)	0.485 (0.628)	0.535 (0.593)
Game theory knowledge	0.359 (0.730)	0.381 (0.703)	0.381 (0.703)	0.000 (1.00)

*Self-assessed degree of risk aversion: *Using a 0-10 scale how much do you like to take risk (high return/high risk investment, extreme sports, lotteries and other risky behavior) or try to avoid it - 0 (I try to avoid it).. 10 (I like risk).* z-stat (with p-values in parentheses) from a non-parametric, two-sided, Mann-Whitney rank-sum test. P-values in round brackets.

Table 2. Average trustor giving according to treatment design and game theory knowledge

No.	Treatments	Mean	St.dev.
1	**TNC** (Trust game)	3.87	2.56
2	**TC_00** (Trust game + corruption with 0 prob. of audit)	5.6	3.11
3	**TC_10** (Trust game + corruption with 10 prob. of audit)	4.53	2.85
4	**TC_50** (Trust game + corruption with 50 prob. of audit)	4.67	2.50
5	Average trustor giving in corruption treatments	4.93	2.81
6	Average trustor giving in corruption treatments (for those who do not know game theory)	5.52	2.79
7	Average trustor giving in corruption treatments (for those who know game theory)	3.34	2.67
8	Average trustor giving in no corruption treatments (for those who know game theory)	2.12	3.19
9	Average trustor giving in no corruption treatments (for those who do not know game theory)	4.25	3.2

N. of obs.: 15 per treatment. Two-sided Mann-Whitney rank-sum test (H0: (1) = (5)) z=1.31, p-value=0.19. Two-sided Mann-Whitney rank-sum test (H0: (6) = (7)) z=2.34, p-value=0.019. Two-sided Mann-Whitney rank-sum test (H0: (8) = (9)) z= -0.23, p-value=0.81.

Table 3. Trustors' expectations on the "corruption tax" (third agent's withdrawal)

VARIABLES	(1)	(2)
$Surplus_{ij}$	0.615***	0.641***
	(0.026)	(0.033)
$Surplus_{ij}$ *DTC10	0.057	-0.008
	(0.037)	(0.046)
$Surplus_{ij}$ *DTC50	0.019	-0.035
	(0.037)	(0.043)
$Surplus_{ij}$*DGAME		0.421***
		(0.104)
$Surplus_{ij}$ *DTC10*D1GAME		-0.305***
		(0.117)
$Surplus_{ij}$ *DTC50*D1GAME		-0.486***
		(0.116)
Constant	0.416**	0.416**
	(0.190)	(0.185)
F-stat		
	587.12	311.59
p>F	0.00	0.00
Observations	450	450
R-squared (within)	0.814	0.824
Number of id	45	45

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 4. Trustors' expectations on trustees' strategy

VARIABLES	(1) All sample	(2) Trustors (GT)	(3) Trustors (no GT)	(4) All sample
$TrustorGive_{ij}$	1.232***	1.058***	1.430***	1.430***
	(0.071)	(0.091)	(0.104)	(0.102)
$TrustorGive_{ij}$*DTC00	-0.252**	-0.327**	-0.284**	-0.284**
	(0.100)	(0.139)	(0.139)	(0.136)
$TrustorGive_{ij}$*DTC10	-0.266***	-0.304**	-0.359***	-0.359***
	(0.100)	(0.147)	(0.136)	(0.133)
$TrustorGive_{ij}$*DTC50	0.036	0.458*	-0.181	-0.181
	(0.100)	(0.274)	(0.128)	(0.125)
$TrustorGive_{ij}$*D1GAME				-0.373***
				(0.140)
$TrustorGive_{ij}$*DTC00*D1GAME				-0.043
				(0.200)
$TrustorGive_{ij}$*DTC10*D1GAME				0.056
				(0.204)
$TrustorGive_{ij}$*DTC50*D1GAME				0.639**
				(0.313)
Constant	0.373*	0.630*	0.245	0.373*
	(0.220)	(0.358)	(0.271)	(0.216)
F-stat	249.55	64.90	195.13	131.40
p>F	0.000	0.000	0.000	0.000
Observations	600	200	400	600
R-squared	0.651	0.596	0.687	0.664
Number of id	60	20	40	60

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 5. Trustor giving, expectations and expected return on giving

	Trustor giving	Expected trustee's payback*	Expected corruption tax*	Expected return on giving
NTC				
TC00				
TC10				
TC50				
Players (GT)				
NTC (n=7)	2.12	1.06		6 percent
TC00 (n=6)	4.16	0.60	64 percent	Negative
TC10 (n=5)	2.2	0.73	74 percent	Negative
TC50 (n=1)	4	1.06	100 percent	negative
Players (No GT)				
NTC (n=8)	4.25	1.43		43 percent
TC00 (n=9)	6.55	1.15	64 percent	5.4 percent
TC10 (n=10)	5.7	1.08	64 percent	2.9 percent
TC50 (n=14)	4.71	1.43	64 percent	15.5 percent

Table 6 Trustee's expectations on the "corruption tax" (third agent's withdrawal)

	(1)	(2)	(3)	(4)
		Trustees	Trustees	
VARIABLES	All sample	(GT)	(no GT)	All sample
$Surplus_{ij}$	0.638***	0.608***	0.665***	0.665***
	(0.026)	(0.041)	(0.033)	(0.035)
$Surplus_{ij}$ *DTC10	0.060	0.142**	0.014	0.014
	(0.036)	(0.067)	(0.044)	(0.046)
$Surplus_{ij}$ *DTC50	0.001	-0.004	-0.001	-0.001
	(0.036)	(0.060)	(0.046)	(0.048)
$Surplus_{ij}$*D2GAME				-0.060
				(0.052)
$Surplus_{ij}$ *DTC10*D2GAME				0.131*
				(0.078)
$Surplus_{ij}$ *DTC50*D2GAME			0.003	
			(0.073)	
F-stat	657.19	203.17	470.22	330.13
p>F	0.00	0.00	0.00	0.00
Constant	0.116	0.427	-0.074	0.116
	(0.184)	(0.323)	(0.221)	(0.184)
Observations	450	170	280	450
R-squared	0.831	0.803	0.850	0.832
Number of id.	45	17	28	45

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 7. Trustee's strategy

	(1)	(2)	(3)	(4)	(5)
VARIABLES	All sample	Trustees (GT)	Trustees (No GT)	All sample	All sample
$TrustorGive_{ij}$	0.669***	0.796***	0.605***	0.796***	0.796***
	(0.057)	(0.091)	(0.069)	(0.095)	(0.101)
$TrustorGive_{ij}$*D2NOGAME				-0.191	-0.191
				(0.117)	(0.123)
$TrustorGive_{ij}$*DTC00	0.248***	0.052	0.373***	0.052	
	(0.080)	(0.119)	(0.104)	(0.125)	
$TrustorGive_{ij}$*DTC10	-0.107	-0.545***	0.069	-0.545***	
	(0.080)	(0.137)	(0.096)	(0.143)	
$TrustorGive_{ij}$*DTC10	0.422***	-0.014	0.692***	-0.014	
	(0.080)	(0.123)	(0.100)	(0.129)	
$TrustorGive_{ij}$*DTC00*D2NOGAME				0.320**	
				(0.161)	
$TrustorGive_{ij}$*DTC10*D2NOGAME				0.614***	
				(0.171)	
$TrustorGive_{ij}$*DTC50*D2NOGAME				0.705***	
				(0.162)	
$TrustorGive_{ij}$*CORRUPTION					0.356***
					(0.083)
$TrustorGive_{ij}$*CORRUPTION*D2GAME					-0.467***
					(0.141)
Constant	0.612***	0.612**	0.612***	0.612***	0.612***
	(0.176)	(0.269)	(0.220)	(0.171)	(0.180)
F-stat	217.19	73.13	164.46	119.73	200.91
	0.000	0.000	0.000	0.000	0.000
Observations	600	220	380	600	600
R-squared	0.618	0.601	0.661	0.643	0.600
Number of id.	60	22	38	60	60

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 8. Aggregate (trustors and trustees) expectations on the corruption tax

VARIABLES	(1)	(2)
$Surplus_{ij}$	0.627***	0.652***
	(0.018)	(0.024)
$Surplus_{ij}$ *DTC10	0.058**	0.005
	(0.026)	(0.033)
$Surplus_{ij}$ *DTC50	0.010	-0.024
	(0.026)	(0.032)
$Surplus_{ij}$*D3GAME		0.035
		(0.043)
$Surplus_{ij}$ *DTC10*D3GAME		0.057
		(0.059)
$Surplus_{ij}$ *DTC50*D3GAME		-0.094*
		(0.057)
Constant	0.266**	0.266**
	(0.132)	(0.132)
F-stat	1245.26	628.31
p>F	0.000	0.000
Observations	900	900
R-squared	0.822	0.824
Number of id.	90	90

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 9. Third agent behavior (corruption tax)

VARIABLES	(1)	(2)
$Surplus_{ij}$	0.708***	0.638***
	(0.026)	(0.037)
$Surplus_{ij}$*D3GAME		0.131***
		(0.050)
$Surplus_{ij}$ *DTC10	0.055	0.083*
	(0.036)	(0.048)
$Surplus_{ij}$ *DTC50	-0.057	-0.068
	(0.036)	(0.050)
$Surplus_{ij}$ *DTC10*D3GAME		-0.006
		(0.073)
$Surplus_{ij}$ *DTC50*D3GAME		0.041
		(0.071)
Constant	-0.201	-0.201
	(0.184)	(0.180)
F-stat	759.04	403.48
p>F	0.000	0.000
Observations	450	450
R-squared	0.850	0.859
Number of id.	45	45

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

References

Berg, J., Dickhaut, J., McCabe, K., (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior* 10(1), 122-142.

Bauman, Y., Rose, E., (2011). Selection or indoctrination: Why do economics students donate less than the rest?. *Journal of Economic Behavior & Organization* 79(3), 318-327.

Anderson, C. J., Tverdova, J. (2003). Corruption, Political Allegiances, and Attitudes toward Government in Contemporary Democracies. *American Journal of Political Science* 47(1), 91-109.

Anokhin, S., Schulze, W., (2009). Entrepreneurship, innovation, and corruption, *Journal of Business Venturing* 24, 465–476.

Arrow, K. (1974), *The Limits of Organisations*, New York, W.W. Norton.

Bacharach, M., Guerra, G., Zizzo, D.J., (2007). The self-fulfilling property of trust: An experimental study. *Theory and Decision* 63(4), 349-388.

Banerjee, R., (2016). Corruption, norm violation and decay in social capital. *Journal of Public Economics* 137, 14-27.

Battigalli, P., Dufwenberg, M., (2007). Guilt in games. *American Economic Review* 97(2), 170-176.

Pranab, B., (1997). Corruption and Development: A Review of Issues. *Journal of Economic Literature* 35(3), 1320–1346.

Bjornskov, C., (2011). Combating Corruption: On the Interplay between Institutional Quality and Social Trust. *Journal of Law and Economics* 54(1), pp. 135 – 159.

Chang, E., Chu, Y., (2006). Corruption and Trust: Exceptionalism in Asian Democracies. *Journal of Politics* 68(2), 259–271.

Della Porta, D., (2000) "Social capital, beliefs in government and political corruption," in S. J. Pharr and R. D. Putnam, eds., *Disaffected democracies: What's troubling the trilateral countries?*, Princeton, NJ: Princeton University Press, 2000.

Dufwenberg, M., (2002). Marital investment, time consistency and emotions. Journal of Economic Behavior and Organization 48, 57-69.

Dufwenberg, M., Gneezy, U., (2000). Measuring beliefs in an experimental lost wallet game. Games and Economic Behavior 30(2), 163-182.

Fischbacher, U., (2007). z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics, 10(2), 171–178.

Guerra, G., Zizzo, D.J., (2004). Trust responsiveness and beliefs. *Journal of Economic Behavior and Organization* 55(1), 25-30.

Guiso, L., Sapienza, P., Zingales, L., (2004). The Role of Social Capital in Financial Development. *American Economic Review*, 94, 526-556.

Guiso, L., Sapienza, P., Zingales, L., (2009). Cultural Biases in Economic Exchange. *Quarterly Journal of Economics* 124, 1095-1131.

Helliwell, J., Huang, H., Wang, S., (2016). New Evidence on Trust and Well-being. NBER Working Paper No. 22450, July 2016.

Helliwell, J., Aknin, L., Shiplett, H., Huang, H., Wang, S., (2017). Social Capital and Prosocial Behaviour as Sources of Well-Being. NBER Working Paper No. 23761, August 2017.

Innes, R., Mitra, A., (2013). Is Dishonesty Contagious? *Economic Inquiry* 51(1), 722–734.

Kaufmann, D., Kraay, A., (2003). Governance and growth: which causes which? The World Bank Working papers. World Bank, Washington, DC.

Knack, S., Keefer, P., (1997). Does social capital have an economic payoff? A cross-country investigation. *Quarterly Journal of Economics*, 112, 1251-1288.

Lambsdorff, G.J., (2003). How corruption affects productivity. *Kyklos* (56), 457–474.

La Porta, R., Lopez de Silanes, F., Shleifer, A., Vishny, R., (1997). Trust in Large Organizations. *American Economic Review* 87(2), 333–38.

Li, H., Xu, L.C., Zou, H., (2000). Corruption, income distribution, and growth. *Economics and Politics* 12, 155–182.

Mauro, P., (1995). Corruption and growth. *Quarterly Journal of Economics* 10(3), 681–712.

Mill, J.S. (1848), *Principles of Political Economy*, London: John W. Parker.

Moreno, A., (2002). Corruption and democracy: A cultural assessment. *Comparative Sociology*, 1(3-4), 495–507.

Pelligra, V. (2005)., Under trusting eyes: The responsive nature of trust, in R. Sugden, B. Gui (Eds.) *Economics and Social Interaction: Accounting for the Interpersonal Relations*. Cambridge: Cambridge University Press.

Pelligra, V., (2010). Trust responsiveness: On the dynamics of fiduciary interactions. *Journal of Socio-Economics* 39(6), 653-660.

Rivera-Batiz, F.L., (2002). Democracy, governance, and economic growth: theory and evidence. *Review of Development Economics* 6(2), 225–247.

Rotondi, V., Stanca, L., (2015). The Effect of Particularism on Corruption: Theory and Empirical Evidence. *Journal of Economic Psychology* 51, 219-235.

Seligson, M. (2001). The Impact of Corruption on Regime Legitimacy: A Comparative Study of Four Latin American Countries. *Journal of Politics* 64(2), 408–433.

Uslaner, E.M., (2008). *Corruption, inequality, and the rule of law*. New York, Cambridge University Press.

Appendix A. Experimental Instructions.

[As follows, we present instructions of TC_50. Original instructions were written in Italian. Subjects received Instructions of each phase only at the end of the previous phase. The only difference between TC_10 and TC_50 concerns the probability of auditing. The difference between TC_50 and TC concerns the fact that, in the latter, there was no audit procedure. Finally, the difference between TC_50 and TNC concerns the fact that, in the latter, C could not make any choice and, therefore, her choice could not be audited.]

Instructions

- Welcome and thank you for participating in this experiment.

- During the experimental session, it is not allowed to communicate with the other subjects. If you have any question, please rise your hand and one of the assistants will come at your place and answer you.

- Following these instructions carefully, you could gain an amount of money (in euro) which depends on both your choices and those taken by the subjects you will interact with. The following rules are the same for all the subjects involved in this experiment.

- During the experimental session, earnings will be expressed in tokens. At the end of the experiment, the overall tokens will be converted in Euro at the exchange rate of 2 tokens = 1 Euro. Your final earnings in Euro will be paid in cash at the end of the session.

General Rules

- 15 subjects participate in this experimental session.

- At the beginning of the experiment, the computer will form 5 groups of 3 subjects each randomly and anonymously.

- During the experiment, each subject will interact with the remaining 2 subjects in her/his group, only. Choices will remain anonymous throughout the experiment.

The Interaction Situation

- The interaction across the 3 group members proceeds in 3 consecutive phases. The final earnings of each subject depend on the choices made by the group members in the 3 phases.

- Schematically:

  - In PHASE 1: the 3 group members will compete in an ability task. The 2 best performers in the group will be assigned to roles A and B. The subject with the lowest score will be assigned to role C.

- In PHASE 2: Both A and B will be assigned an endowment of 10 tokens. A will choose how many tokens of her/his endowment to send to B. B will receive a number of tokens that is equal to three times those sent by A. Of the received tokens, B will choose how many tokens to send to A.

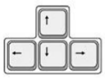
- In PHASE 3: C will be assigned an endowment of 10 tokens. C will choose how many tokens of the “surplus” generated during the interaction between A and B in PHASE 2 to keep for herself/himself. The “surplus” is given by 2 times the number of tokens sent by A in PHASE 2. Given the choice of C, earnings obtained by A and B in PHASE 2 will be reduced accordingly, in proportion of the share of surplus obtained by each subject in PHASE 2. With a probability of 50%, the choice of C will be audited by the computer. If C has not kept any token of the surplus, the audit procedure does not exert any effect on her/his earnings. Instead, in case of auditing and if C has kept a positive number of tokens, then her/his earnings will be reduced by the number of tokens kept plus a fine that is equal to 1 token for every 2 subtracted from the surplus. The audit procedure does not exert any effect on A’s and B’s payoffs that will remain equal to those obtained in PHASE 2 minus the tokens kept by C in PHASE 3.

- The instructions of each of the three phases will be distributed at the beginning of the corresponding phase.

PHASE 1: the Ability Competition

- In PHASE 1, your and the other two members of you group will compete in an ability task.

- The computer will show a screen containing a number of “sliders” of the following form:

- By using the mouse and the arrow keys , your task will be to align the cursor on the left of the slider to the value “50.”

- The competition lasts 120 seconds. At the end of the 120 seconds, the computer will inform you about your final score given by the number of sliders for which you successfully centered the cursor on the value “50.”

- Given the final scores, each group member will be assigned to one of three possible roles, either A, B, or C. In particular, the two subjects with the highest scores will be randomly assigned either A or B. Instead, the subject totalizing the lowest score will be assigned role C. Ties will be randomly broken by the computer.

PHASE 2: A's and B's Choices

- In PHASE 2, only A and B will make choices. C will not make any choice in this phase.
- Both A and B will be assigned an endowment of 10 tokens.
- A will choose how many tokens to send to B. She/he can choose to send any number of tokens that is included between 0 and 10.
- The amount sent by A will be tripled such that B will receive 3 tokens for each sent by A.
- Given the received amount, B will choose how many tokens to send to A. She/he can choose to send any number of tokens that is included between 0 and 3 times the amount initially sent by A.
- Given their choices, A's and B's earnings in PHASE 2 will be given by:
  - For A: 10 tokens – tokens sent to B + tokens sent by B;
  - For B: 10 tokens + 3*tokens sent by A – tokens sent to A.
- Example: If A sends x tokens to B and B sends y tokens to A, A earns 10 − x + y tokens and B earns 10 + 3x – y tokens.
- C's earnings are null in PHASE 2.
- B will make her/his choice before knowing the number of tokens effectively sent by A. In particular, B will choose how many tokens to send to A for each possible amount that A could have sent to her/him (1, 2,..., 10 tokens). Since there are 10 possible cases, B will make 10 choices.
- Given the 10 choices made by B, only the one corresponding to the effective choice made by A will be used to determine earnings in PHASE 2.
- At the end of PHASE 2, A and B will be informed about their earnings in tokens before C makes her/is choice.

PHASE 3: the Choice of C

- In PHASE 3, only C will make her/his choice. A and B will not make any choice in this phase.
- C will be assigned an endowment of 10 tokens.
- C will choose how many tokens of the "surplus" generated during the interaction between A and B in PHASE 2 to keep for herself/himself. The "surplus" is given by 2 times the number of tokens sent by A in PHASE 2. Thus, the "surplus" increases in the number of tokens sent by A in PHASE 2 and is null if A has sent nothing to B.

- Given the choice made by C, earnings obtained by A and B in PHASE 2 will be reduced accordingly, in proportion of the share of "surplus" obtained by each subject in PHASE 2. In particular, on the basis of the "surplus" generated during the interaction between A and B in PHASE 2, the shares obtained by A and B are given by:

$$Surplus = 2 * tokens\ sent\ by\ A\ to\ B$$

$$Share\ of\ A = \frac{Tokens\ of\ A\ at\ the\ end\ of\ PHASE\ 2 - Endowment\ of\ 10\ Tokens}{Surplus}$$

$$Share\ of\ B = \frac{Tokens\ of\ B\ at\ the\ end\ of\ PHASE\ 2 - Endowment\ of\ 10\ Tokens}{Surplus}$$

  The reduction in A's earnings will be equal to the "Share of A" multiplied by the choice of C, while the reduction in B's earnings will be equal to the "Share of B" multiplied by the choice of C. In case of a negative share assigned to a subject, her/his earnings will not be influenced by the choice of C. Instead, if C does not keep any token for herself/himself, A's and B's earnings in PHASE 2 will remain unchanged.

- Example. Suppose that, at the end of PHASE 2, the "surplus generated during the interaction between A and B in PHASE 2 is equal to 10 tokens, A's earnings are equal to 16 tokens and, finally, B's earnings are equal to 14 tokens. Suppose that C chooses to keep 10 tokens for herself/himself. Given the previous expressions, A's and B's shares of "surplus" are given by $Share\ of\ A = \frac{16-10}{10} = 0{,}6$ e $Share\ of\ B = \frac{14-10}{10} = 0{,}4$. Thus, the reduction in earnings will be, respectively, $Share\ of\ A * Choice\ of\ C = 0{,}6 * 10 = 6$ tokens for A and $Share\ of\ B * Choice\ of\ C = 0{,}4 * 10 = 4$ tokens for B.

- C will make her/his choice before knowing the effective size of the "surplus" generated in PHASE 2. In particular, C will choose how many tokens to keep for herself/himself for each possible amount of the "surplus" (2, 4,..., 20 tokens). Since there are 10 possible cases, C will make 10 choices.

- Given the 10 choices made by C, only the one corresponding to the effective size of the "surplus" generated in PHASE 2 will be used to determine earnings.

- With certain probability, C's choice can be audited by the computer.

- If C's choice is not audited, C's earnings will be equal to the endowment of 10 tokens plus the tokens of the "surplus" generated in PHASE 2 kept by C for herself/himself.

- Instead, in case of auditing and if C has kept a positive number of tokens of the "surplus" generated in PHASE 2 for herself/himself, then her/his earnings will be reduced by the number of tokens kept plus a fine that is equal to 1 token for every 2 subtracted from the surplus. In this case, C's earnings will be equal to the endowment of 10 tokens minus one token for every two subtracted from the surplus.

- The audit procedure does not exert any effect on A's and B's payoffs that will remain equal to those obtained in PHASE 2 minus the tokens kept by C in PHASE 3.

- Example: Suppose that C keeps z tokens from the effective "surplus" generated in PHASE 2. Then, C earns 10 + z tokens if her/his choice is not audited, while she/he earns 10 – z/2 tokens if her/his choice is audited.

- The audit procedure follows a random scheme. In particular, after C makes her/his choice, the computer will select an integer included between 1 and 100 randomly and with equal probability. If the selected number is included between 1 and 50, then C's choice will be audited. Instead, if the selected number in included between 51 and 100, then C's choice will not be audited. This means that C's choice will be audited with a probability of 50 percent. Moreover, the random selection of the number neither depends on the choice of C, nor on those made by the other group members.

- At the end of PHASE 3, A and B will be informed about how many tokens of the "surplus" have been kept by C for herself/himself, while C will be informed about the outcome of the audit procedure and her/his earnings in tokens.

A's and B's conjectures

- Before being informed about the final results of the experiment, A and B will be given the opportunity to increase their payoffs by guessing the choices of the other group members in PHASE 2 and PHASE 3. Instead, C will not express any conjecture. The procedure used to pay conjectures is such that A and B should express their conjecture accurately and truthfully. Indeed, A and B will receive 3 tokens only if their conjectures will be correct.

B guesses A's choice

- B has to guess the number of tokens sent by A in PHASE 2.

- If B's conjecture is correct, B will earn 3 additional tokens.

A guesses B's choices

- In PHASE 2, B has made 10 choices about the tokens sent by A, one for each of the possible positive amounts that A could send (1, 2,..., 10 tokens). A has to guess the number of tokens chosen by B for each of the 10 possible cases.

- The computer will randomly select one of the 10 conjectures expressed by A. If the selected conjecture is correct, then A will earn 3 additional tokens.

A and B guess C's choices

- In PHASE C, C has made 10 choices about the tokens to keep from the "surplus" generated during the interaction between A and B in PHASE 2, one for each of possible positive size of the "surplus" (2, 4,..., 20 tokens). Both A and B have to guess the number of tokens kept by C for each of the 10 possible cases.

- For each of the two subjects, A and B, the computer will randomly select one of the 10 conjectures. If the selected conjecture is correct, then the corresponding subject will earn 3 additional tokens.

- After expressing their conjectures, A and B will be informed about their correctness and final earnings from participating in the experiment.