The effects of personality, risk and other-regarding attitudes on trust and reciprocity

Gerardo Sabater-Grande[1], Aurora García-Gallego[1], Nikolaos Georgantzís[1,2] and Noemí Herranz-Zarzoso[3]

[1] LEE & Department of Economics, Universitat Jaume I, Castellón, Spain

[2] CEREN EA 7477, Burgundy School of Business, Université Bourgogne Franche-Comté, Dijon, France

[3] Department of Economic Analysis, Universitat de València, Valencia, Spain.

*** Correspondence:**

Nikolaos Georgantzís. E-mail: mail: nikolaos.georgantzis@bsb-education.com

Postal address: 29 Rue Sambin, 21000 Dijon, France.

Abstract

This paper reports experimental results on the determinants of trust and reciprocity in the context of a genuinely sequential, binary Trust Game. Apart from behavior in the main experiment, subjects' risk attitudes and inequality aversion are elicited, as well as the traits of neuroticism and agreeableness, captured through the five-factor model. The findings suggest that trustors' (first movers) behavior is affected by their loss aversion, while trustees' (second movers) reciprocal behavior is not explained by any of their other-regarding attitudes, but, rather, by their agreeableness.

Keywords: Behavioral economics, inequality attitude, risk attitude, trust, personality.

JEL codes: C9, D91

1. Introduction

The Trust Game (TG, henceforth), originally developed by Berg *et al*. (1995) is a game designed to measure trust and reciprocity. In the original version of the game both trustor (or investor) and trustee receive an equal endowment. The investor first decides how much of her endowment to transfer to the trustee. That amount arrives to the trustee multiplied by some factor and then, the trustee chooses to send all, some, or none of the received money back to the investor. Alternative versions of the TG have been implemented in the lab, including a continuous investment game (Costa-Gomes *et al*., 2014) and a discrete binary game version (Gambetta, 1988). Although the subgame perfect equilibrium of the TG with selfish players involves no trust and no reciprocity, experimental studies have typically found positive amounts invested by first movers and reciprocal behavior by second movers (Johnson & Mislin, 2011). To explain the deviation of observed behavior from the selfish subgame perfect equilibrium prediction, it suffices to assume that second movers, for some reason, are reciprocal to the trust received and this is anticipated by first movers. In such cases, the trust decision becomes similar to a choice under uncertainty in the sense that the level of trust by first movers may depend on their risk attitudes, while the reciprocity of second movers depends on their other-regarding attitudes and, potentially, on their personality.

The present paper connects three separate lines of research. First, the literature that links risk preferences with trust. Since both risk and trust are associated with unforeseen future events, trust can be related to risk attitudes by means of the expected monetary returns of trusting behavior. As noticed by Eckel & Wilson (2007) "the decision by someone whether to trust another person is viewed as similar to placing a risky bet on the trustworthiness of an anonymous counterpart in a situation where both can gain from reciprocal exchange". Such a statement, however, was either rejected or not systematically supported by several related studies such as Houser *et al*. (2010), Ashraf *et al*. (2006), Garapin *et al*. (2015) or even Eckel & Wilson (2007). The hypothesis that financial losses are substantially different from the regret felt when being betrayed by a person you have trusted, has been formally addressed in a series of studies in which the social component is eliminated, getting puzzling

results.[1] Exceptions to this are the works by Schechter (2007), Sapienza *et al*. (2013) and Chetty *et al*. (2021), who confirm that trust is negatively related with risk aversion, while Beattie (2019) reports that the relation is not robust to different definitions of the amount sent variable and specifications of the risk model. Finally, Saadoui *et al*. (2019) find the opposite results and suggest that risk seekers exhibit less trust. The authors explain this result through the role played by emotions: negative emotions influence trust positively and people under negative emotional state are found to be less risk seeking.

The second line of research explores the link between inequality aversion and behavior in the TG. Anderson *et al*. (2006) find that, while allocators' behavior is not affected by inequality at all, paying subjects different show-up fees reduce first movers' trust, when payments are made privately. In a similar unequal-reward design, Lei & Vesely (2010) find that rich investors trust their in-group members significantly, that poor first movers do not exhibit such in-group bias, and that second movers' reciprocation does not discriminate among rich or poor trustors. Brülhart & Usunier (2012) find no evidence of a relation between investor transfers and trustees' "wealth". Xiao & Bichieri (2010) compare trustees' behavior under equal and unequal endowments (more specifically, the trustee's endowment is lower than that of the investor) and obtain that more than half of the trustees do not reciprocate in the cases in which this would increase disadvantageous inequality. Hargreaves *et al*. (2013) report that neither inequality nor the knowledge of it have any effect on investors' behavior, whereas common knowledge observed inequality reduces the average return rate from trustees. Finally, Rodriguez-Lara (2018) shows that, regardless of the first mover's endowment, trustees return the same proportion of funds. Behavioral spillovers across different games eliciting other-regarding preferences are also relevant for our study. Cox (2004) provides evidence of such spillovers across a dictator and a TG, concluding that reciprocal transfers are associated with other-regarding preferences. Ashraf *et al*. (2006) suggest that transfers in the TG are significantly correlated to transfers in a modified dictator game. Garapin *et al*. (2015) find that pro-social concerns elicited through the ring test (Griesinger & Livingston, 1973)[2] are not a strong predictor of behavior in the TG.

[1] Fetchenhauer & Dunning (2012) find higher trust when playing against computers than with human opponents. In the same vein, results by Bohnet & Zeckhauser (2004) or Aimone *et al*. (2014) suggest that trust is increased when betrayal by a human is not possible.

[2] This test determines a social value orientation for each subject along a spectrum ranging from altruistic to aggressive.

The third line of research links personality to trust and reciprocity. Several studies have tried to link trust and personality traits, using a wide array of personality measures. Ibáñez *et al*. (2016) find that trust behavior associates to the facets of extraversion's warmth and impulsivity of positive urgency. Authors present some physiological results showing that first movers scoring high in psychopathy exhibited increased electro-dermal activity and reduced evoked heart rate deceleration at the moment in which they were asked to decide whether to trust. Regarding second movers, lack of reciprocity was found to relate with a psychopathic, highly disinhibited and impulsive personality. As regards this last finding, four studies relate Machiavellianism with behavior in the TG. In particular, Gunnthorsdottir *et al*. (2002) and Szijjártó *et al*. (2018) find no correlation between Machiavellianism scores and trust, whereas individuals scoring high in Machiavellianism are less trustworthy. Contrasting with these findings, Burks *et al*. (2003) obtain that a high score in Machiavellianism predicts distrust, but not a lack of trustworthiness. This last finding is supported by Ben-Ner & Halldorsson (2010) in a later study. Thielmann & Hilbig (2015) investigated the relation between trustworthiness and the HEXACO personality factors and found that trustworthiness is mainly driven by unconditional kindness.

The Big Five Personality Model is arguably one of the most accepted personality theories, using five basic dimensions: extraversion (E), agreeableness (A), conscientiousness (C), neuroticism (N), and openness to experience (O). These five dimensions describe personality traits (John *et al*., 2008). Using the Catell's 16 PF-R test, Fahr & Irlenbusch (2008) find that low-anxiety individuals trust and reciprocate more than high anxiety ones. Using the NEO-FFI test (Costa & McCrae, 1992), Ben-Ner & Halldorsson (2010) suggest that a substantial portion of the amounts sent in the TG, more than just trusting, reflects unconditional kindness towards others. Alarcon *et al*. (2018) find that personality traits predict beliefs and intentions but not trust itself. Contrary to this result and using the NEO-PI, Swope *et al.* (2008) find that extraversion is related with trust. Using a revised version of the same test, the results of Johnson & Mislin (2011) show personality facets can help in explaining decision making when subjects are in a weak situation (investors), but not in situations in which they are in a strong position (allocators). Specifically, their results are compatible with the hypothesis that neuroticism influences trust negatively, and are incompatible with the hypothesis that agreeableness affects trust positively..

Using different risk and other-regarding preference elicitation tasks, our study addresses, on one hand, the links among risk, other-regarding attitudes and personality, and the links between trust and

reciprocity in the TG on the other. Assuming that trust is a risky decision, Thielmann & Hilbig (2015a) distinguish three core components in trust behavior: attitudes towards risky prospects, trustworthiness expectations and betrayal sensitivity. Based on these three underlying mechanisms, our study addresses three differentiated hypotheses for trustors. First, since trust among strangers (McKnight *et al*., 1998) is not based on information about the trustee's intention, this inability to predict the trustee's behavior is related to the positive probability of making losses as the effect of trusting the partner (Rousseau *et al*., 1998). Consequently, the first hypothesis to study is whether:

Hypothesis 1 (H1): More risk averse trustors are less likely to trust second movers.

In situations of trust, in order to form an expectation about the other's behavior, the trustor can consider different sources of information, such as trust cues (i.e., reputation), prior trust experiences, or social projection (Thielmann & Hilbig, 2015b). In the last option, trustworthiness expectations are created by projecting own cooperativeness onto others (Thielmann & Hilbig, 2014). This social projection mechanism implies that personality traits driving cooperation should be associated not only with trustworthiness but, also, with trusting behavior (Krueger *et al*., 2008). Since agreeableness has been identified as the main personality facet predicting cooperation (Denissen & Penke, 2008), we hypothesize that:

Hypothesis 2 (H2): Agreeableness favors trust by first movers.

Trusting is necessarily associated with vulnerability towards the trustee's actions (e.g., Mayer *et al.*, 1995; Rousseau *et al.*, 1998). Specifically, Lee & Selart (2015) argue that subjects scoring high in neuroticism are particularly vulnerable to a betrayal experience and will adopt a more skeptical relationship with strangers. Based on this, we build our third hypothesis:

Hypothesis 3 (H3): Neuroticism hinders trust.

Following Ozer & Zheng (2017), trustworthiness "is unconditional kindness generated by social norms or values that an individual adheres to". Deviating from these social norms causes psychological cost for the individual that considers her "self" in relation to "others". This potential prosocial behavior can be explained through attitudes towards inequality aversion, and it is expected that:

Hypothesis 4 (H4): Trustees with higher levels of inequality aversion reciprocate more.

In terms of personality traits and additionally to the aforementioned predictive role of agreeableness on cooperation, Thielmann & Hilbig (2015b) state that unconditional kindness is an underlying determinant of trustworthiness. Given that unconditional kindness is a component of the Big Five-Agreeableness factor, we hypothesize that:

Hypothesis 5 (H5): Agreeableness favors reciprocity.

2. Materials and methods

The experiment[3] was run at the economics experimental lab (Laboratorio de Economía Experimental, LEE) of Universitat Jaume I (UJI) in Castellón (Spain). Prior to the experimental sessions, informed consent was obtained from all participants acknowledging acceptance of our data management protocols with respect to anonymity, confidentiality and exclusive use for scientific research. The tool ORSEE32 was used to recruit 220 undergraduate students from different degrees (psychology, economics, business administration, law, etc.) at UJI. The average duration of a session was approximately one hour, and the average payoff was €15. In the main experiment, the sample was divided into 110 trustors (71 females) and 110 trustees (67 females).

At the beginning of the session, subjects were given written instructions, also read aloud by the experimenters, with specific questions being answered privately. Participants performed several tests before playing the TG, so that personality, risk attitudes as well as attitudes towards inequality could be elicited. Subjects' risk preferences and their attitudes towards (dis)advantageous inequality were elicited using real monetary incentives.[4] In the following, the measures used in the experiment are explained in detail.

---

[3] Our data are part of a larger database obtained from sessions that took place some years earlier. See Ibáñez *et al*. (2016) and Sabater-Grande *et al*. (2020) for more details about the experimental design.

[4] The empirical evidence reported here is part of a larger study on personality traits and behavior in a series of games like prisoners' dilemma, ultimatum, trust, dictator games, as well as risk preferences and inequality aversion tests. In order to avoid wealth accumulation effects and portfolio strategies, payment was contingent on performance in one of the games, chosen randomly at the end of the session,

2.1. Personality

The McCrae and Costa's Five Factor Model is used for measuring different aspects of personality. The Spanish NEO-PI-R by Costa & McCrae (1999) is a 240-item self-report measure that quantifies 30 specific traits (facets) that define the five personality factors or domains: N, E, O, A, and C. Items are responded to on 5-point Likert scales ranging from 0 (strongly disagree) to 4 (strongly agree). In general terms, the research questions formulated in the previous section address our interest mainly to the domains of agreeableness (A) and neuroticism (N).

2.2. Elicitation of Risk Attitudes

In order to elicit subjects' risk attitudes, we use the Sabater-Grande & Georgantzís's (2002) lottery-panel task (hereafter SGG).[5] The lottery-panel task is displayed in Table 1. In this task, subjects face eight subtasks called panels 1, 2, 3, …, 8, presented in this order on a sequence of decision screens. Each panel corresponds to ten lotteries, from which subjects have to choose their preferred one. The task joins two sub-tasks: the choices in panels 1-4 and the choices in panels 5-8. In panels 1 to 4, each lottery is defined by the probability of winning a prize of €X (else nothing) at no cost, but a degenerate lottery of a certain (Prob=100%) reward (€1) is also included. Therefore, the unfavorable options of all other lotteries (€0) can be considered as a loss with respect to the €1 sure alternative. Compared to the first 4 panels, panels 5 to 8 set the certain option to X=€0, ruling out the possibility of loss considerations. By inspection, the farther right a subject chooses, the less risk averse she is. For comparability of choices across panels, our data analysis includes just the winning probability of the lottery chosen. To calculate earnings from the SGG lottery-panel task, one of the eight panels is randomly picked and the lottery preferred by each subject in that panel is executed.

In all panels, the winning probability varies from Prob=0.1 to Prob=1 in steps of 0.1. Prizes are designed such that within a panel, the expected value of the lotteries increases linearly with the probability of not winning by a constant *t* over a fixed gain of €1 in panels 1–4, and €0 in panels 5–8. Thus, *t* is a panel-specific risk premium parameter, which generates an increase in the expected value of the lotteries as we move from safer to riskier options within the same panel. This parameter is increased from panel 1 to 4, and from panel 5 to 8. Intuitively, a subject should be expected to make riskier choices when moving from panel 1 to panel 4, as well as from panel 5 to panel 8. In order to

[5] A detailed description of this test may be found in Barreda-Tarrazona *et al*. (2020). It is also discussed in association with other risk elicitation tasks by Attanasi *et al*. (2018).

estimate the participants' score in risk attitudes, an exploratory factor analysis with principal axes factor analysis and varimax rotation was performed. According to eigenvalue and principal factor analysis, two factors emerged: Factor 1 (risk aversion, F1), comprising Panels 5–8 (with factor loadings from 0.70 to 0.87); and a relatively independent (Factor correlation = 0.20) Factor 2 (risk aversion F2) comprising Panels 1–4 (with factor loadings from 0.73 to 0.83). These two factors actually explain 65.5% of the variance.

Given that the sure option in panels 1 to 4 is 1€, the unfavorable outcome represents a loss with respect to the sure option reference point. Moreover, panels 5 to 8 have a reference point of 0€ with respect to which the unfavorable option represents no loss. Therefore, panels 1 to 4 may be interpreted as a composite measure of risk and loss aversion, while panels 5 to 8 may be considered to represent risk aversion alone. Consequently, the joint use of the two risk subtasks is used to address the extra role of loss aversion (risk aversion for gains and risk seeking for losses).

2.3. Elicitation of Inequality Attitudes

For the elicitation of the Advantageous Inequality Aversion (AIA) and Disadvantageous Inequality Aversion (DIA), subjects perform two tasks related to the dictator game. In each task, participants play both the dictator's and the recipient's roles, and one of the roles is randomly chosen to determine the subjects' final payoff. The underlying assumption in both tasks is that the dictator's preferences include two components, one for own and one for the partner's gains. Due to a qualitative difference between advantageous and disadvantageous inequality, two different tasks are used in order to address the relative weight of each player's other-regarding component in each context. With the aim of not inducing a different perception of the two tasks due to format heterogeneity, we modified the advantageous inequality task in order to create the disadvantageous inequality tak (see Table 2 and Table 3).

In the version of the task for measuring AIA displayed in Table 2, the dictator has to decide which part of the payoff of €15 (if any) is willing to give to the partner, achieving a more equitable payoffs' distribution. More specifically, in this AIA-task, subjects are given two lists of 16 payoff vectors where the first (second) element of each payoff vector corresponds to the dictator's (recipient's) payoff.

Observe in Table 2 that the left-hand side payoff vector contains equal payoffs for both subjects varying from (€15, €15), (€14, €14) up to (€0, €0), whereas the right-hand side payoff vector is always (€15, €0). Subjects were asked to choose at which point, if any, they preferred to switch from the left to the right payoff vector. For the sake of consistency, participants were not allowed to switch back and forth. Specifically, for this task, a switching point (SP) value of "0" indicates that the subject prefers the left-hand side payoff to the outcome (€15, €0) in all cases considered. At the other extreme, a SP of "16" denotes a situation in which the subject prefers the outcome (€15, €0) to the egalitarian payoffs in all cases considered. In intermediate positions, a SP of "1" indicates that the subject always prefers the egalitarian payoffs to (€15, €0) except when the outcome is (€0, €0). A SP of "2" denotes a preference for all the egalitarian payoffs to (€15, €0) except for outcomes (€0, €0) and (€1, €1), and so on. We use the SP ranging from 0 to 16 in order to elicit a parameter of aversion to advantageous inequality. Using the utility function of Fehr and Schmidt (1999), like Blanco *et al*. (2011) we calculate point estimates of the subjects' AIA parameter (β). Given that the right-hand payoff vector is (€15, €0), the AIA parameter β is calculated as:

$$\beta = 1 - \frac{SP}{15}$$

The lower the SP, the higher the value of β is, and, therefore, the higher the aversion to advantageous inequality. The SP is used as the advantageous inequality score to carry out the data analysis.

In the version of the task for measuring DIA displayed in Table 3, the left-hand side of the payoff vector contains identical equal payoffs to those contained in the AIA, but the right-hand side always consists of the pair of payoffs (€0, €15). Subjects are asked to choose at which point, if any, they prefer to switch from the left to the right-hand side pair of payoffs. For this task, a SP value of "0" would indicate that the subject prefers the egalitarian payoffs to the outcome (€0, €15) in all cases. In the other extreme, a SP of "16" corresponds to a subject who always prefers the outcome (€0, €15) to any of the egalitarian ones. In intermediate positions, a SP of "1" indicates that the subject prefers the egalitarian payoffs to (€0, €15) in all cases except when the outcome is (€0, €0). A SP of "2" denotes a preference for all the egalitarian payoffs to (€0, €15) except for outcomes (€0, €0) and (€1, €1), and so on and so forth. It is worth mentioning that the only reason for a subject to choose the right-hand side options is a willingness to sacrifice her own payoff to increase the partner's payoff, representing a subject's kind of generosity. We use the SP ranging from 0 to 16 with the aim of calculating the parameter of aversion

to disadvantageous inequality, which we call the DIA parameter (α). Like in the case of β in the AIA, the DIA parameter is calculated as:

$$\alpha = 1 - \frac{SP}{15}$$

Observe that the lower the value of SP, the higher the level of aversion to disadvantageous inequality is, corresponding to a higher value of α. The SP is used as the disadvantageous inequality score to conduct the data analysis.

2.4. Trust Game

The experiment is based on a discrete version of the TG by Berg *et al.* (1995) with binary choices and no particular framing. In this context, half of the participants acted as trustors (Player 1, N = 110), whereas the rest acted as trustees (Player 2, N = 110). Instructions to the subjects never mentioned the words 'trust', 'investment' or 'reciprocity', in order to avoid undesirable experimenter demand effects. Figure 1 presents the tree of the game, with the payoffs and the number of subjects choosing each strategy.

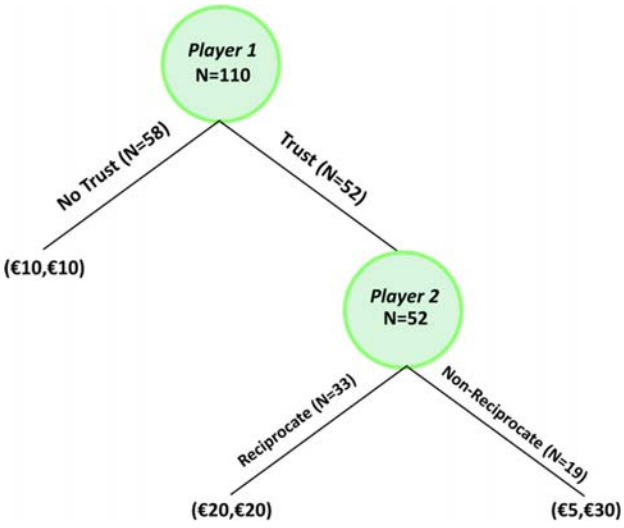


Figure 1: Extensive form of the binary TG.

Observe that if the trustor decides not to trust, each player earns 10€ with certainty. But if the trustor trusts the partner, the latter will have to choose whether to reciprocate raising each players' earnings to 20€-, or to behave individually, raising own payoffs to 30€ and letting the trusting player down (with 5€). Pairs were randomly formed and the game was played once in its genuine sequential form. Each trustor made the decision whether to trust or not before the trustee made the second stage decision, provided that the trustor had decided to trust in the first place. As shown in Figure 1, 52 (35 females) out of 110 first movers decided to trust. From the 52 active second movers, 33 (22 females) reciprocated and 19 (11 females) exploited first movers' trust towards them.

3. Data Analysis and Results

In order to study the relationship among risk attitudes, inequality aversion, as well as personality variables on the TG behaviors, non-parametric median comparison tests and regression analysis have been performed. Like Müller & Schwieren (2020), we only have clear hypotheses regarding the players' behavior in the TG with respect to neuroticism and agreeableness facets. In consequence, other personality traits of the Big Five model that presumably could lead to more or less sending/reciprocation, showed non-significant effects on TG behavior and were excluded from our regressions.

3.1. Descriptive statistics

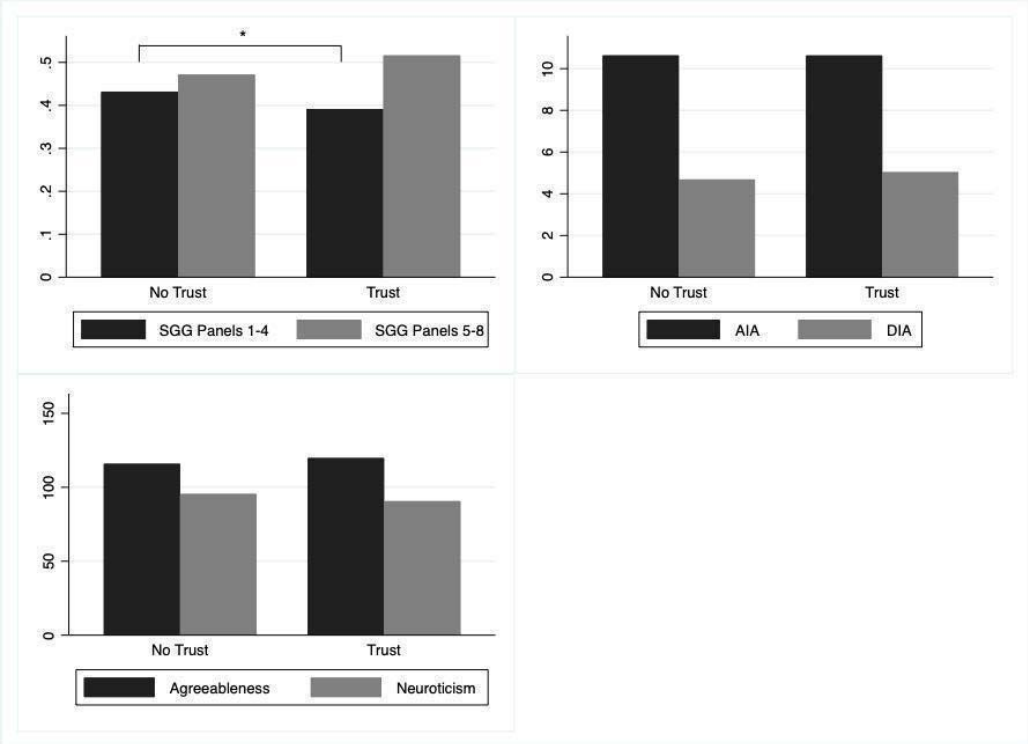
Table 4 presents descriptive statistics (median and standard deviation) of the explanatory variables included in our study. These variables are:

- SGG P1-P4: Per subject mean winning probability of the lotteries chosen in panels 1 to 4. It ranges from 0.1 to 1.
- SGG P5-P8: Per subject mean winning probability of the lotteries chosen in panels 5 to 8. It ranges from 0.1 to 1.
- AIA: switching point ranging from 0 to 16.
- DIA: switching point ranging from 0 to 16.
- Agreeableness (A): score ranging from 0 to 160.
- Neuroticism (N): score ranging from 0 to 160.

Moreover, Shapiro-Wilk (S-W) test p-values are reported in order to test the normality of the data. We observe that all variables, with the exception of neuroticism, are not normally distributed. In consequence, we will use non-parametric tests in order to compare samples.

3.2. Test comparisons

First, we split the sample of trustors according to their strategy. Figure 2 shows that trusting and non-trusting first movers exhibit similar means in all explanatory variables except for the SGG Mean P1-P4. Specifically, using a Mann-Whitney test comparing medians, we find that trustors that actually trust their partner correspond to a significantly lower level of loss aversion compared to the ones that do not trust in the TG.[6]



[6] The p-values for a Mann-Whiney test corresponding to each explanatory variable are 0.100 (SGG P1-P4), 0.301 (SGG P5-P8), 0.781 (AIA), 0.648 (DIA), 0.355 (A), 0.424 (N). Given that the statistical difference obtained is significant at a 10% level and there are 52 trusting first movers in the sample, we check whether such a difference is due to a low statistical power. After conducting an ex-post power analysis using Stata with power set at 0.80 and probability at 0.05, it is obtained that: sample sizes would have to increase to N=303 trusting and N=303 non-trusting first movers in order for the actual difference to reach significance at 5% level.

Figure 2. Mean differences in risk & inequality aversion and personality scores between non-trusting (N=58) and trusting (N=52) first movers. SGG P1-P4: Per subject mean winning probability of the lotteries chosen in panels 1 to 4; SGG P5-P8: Per subject mean winning probability of the lotteries chosen in panels 5 to 8; AIA: advantageous inequality aversion; DIA: disadvantageous inequality aversion. *p < 0.10.

The sample of active deciding (N=52) second movers is now split according to their strategy in the second stage of the game. In Figure 3, we see no differences in risk and inequality preferences between trustees that reciprocate and the ones non-reciprocating.[7] Regarding personality characteristics, it can be observed that trustees who display a reciprocal behavior have significantly higher scores in agreeableness than the ones who have opted for the individualistic reaction to their trusting counterpart.[8]

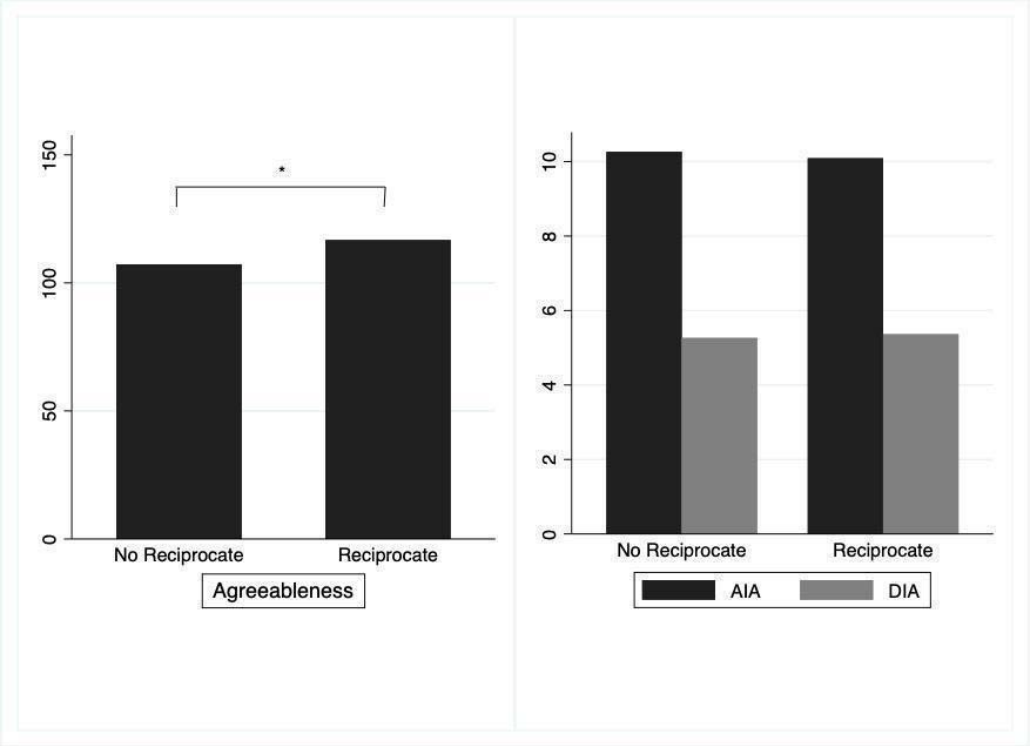


Figure 3. Mean differences in risk & inequality aversion and personality scores between non-reciprocating (N=33) and reciprocating (N=19) trustees. SGG P1-P4: Per subject mean winning probability of the lotteries chosen in panels 1 to 4; SGG P5-P8: Per subject mean winning probability of the lotteries chosen in panels 5 to 8; AIA: advantageous inequality aversion; DIA: disadvantageous inequality aversion. *p < 0.10..

[7] Mann-Whiney p-values corresponding to inequality aversion are 0.969 (for AIA) and 0.969 (for DIA), respectively.

[8] The p-value for a Mann-Whiney test corresponding to agreeableness is 0.064.

3.3. Econometric Analysis

Given the binary nature of our dependent variables, probit regressions have been estimated in order to deeply analyze the effects considered in our research questions. In our data analysis, the joint effect of risk attitudes and inequality aversion on trusting behavior as well as the two personality traits (A and N) that could hypothetically affect in a clear direction decision making in the TG, have been included.

Two probit models, one for each type of player, have been estimated. Results corresponding to the probit estimation of trustors' behavior are reported in Table 5. More specifically, the variables used are:

- Dependent variable: Trust, that takes value 1 when the individual decides to trust, 0 otherwise.
- Independent variables: SGG P1-P4, SGG P5-P8, AIA, A, N and gender (that takes value 1 if the subject is female, 0 otherwise).

The results indicate that trustors' behavior is affected by their risk preferences, which is in line with the first hypothesis. Confirming H1 only in the domain of gains, more loss averse trustors in our study are less likely to trust the partner. However, neither inequality aversion nor personality traits affect trusting behavior, therefore rejecting H2 as well as hypothesis H3. Furthermore, no gender differences are found for first movers' behavior.

In order to account for the effect of inequality aversion and agreeableness on reciprocating behavior, probit regressions are estimated for trustees' behavior. Results are reported in Table 6. The variables used are:

- Dependent variable: Reciprocation, that takes value 1 when the individual decides to reciprocate, 0 otherwise.
- Independent variables: AIA, A and gender.

The results indicate that hypothesis H4 is rejected. In fact, trustees' behavior is not affected by their inequality aversion. However, H5 is confirmed for second movers, suggesting that subjects scoring high in agreeableness are more likely to reciprocate. Lastly, like in the trustors' case, we do not find any gender effect on trustees' behavior.

4. Discussion and main conclusions

We have studied the effect of risk attitudes, other-regarding preferences and personality on behavior in a binary trust game. In line with Nguyen's (2016) results, we argue that the Expected Utility Theory can be inadequate to analyze the link between trust and risk preferences if individuals evaluate risky prospects against a reference point or if they are endowed. In such cases, due to loss aversion, subjects may opt to keep their endowment rather than taking the risk of losing their money. Thus, like Chetty *et al*. (2021), we have avoided risk elicitation procedures that are underpowered. The use of a more sophisticated risk preference elicitation mechanism like the SGG lottery-panel task, where loss aversion is accounted for, allows for finding risk-trust relationships.

Although no other-regarding component nor personality in the behavior of first movers seems to be relevant, trustors' risk attitudes matter. In fact, trustors face the decision whether to trust as a risky choice, anticipating that some reciprocal behavior will reward trust with the egalitarian or higher payoff. Similar to Nguyen (2016), our data show that more loss averse first movers tend to trust less, resorting to the safe outside option. The link between strategic uncertainty and risk preferences may explain the relation between trust and risk attitudes. Depending on the features of the strategic environment analyzed and the risk elicitation procedure used, risk preferences can be related to beliefs about other players' decisions (Brañas-Garza *et al*., 2004). Thus, the trustor's decision can be explained through risk preferences towards lottery choices if the risk elicitation task is sophisticated enough to capture attitudes towards strategic uncertainty.

Regarding second movers, we find that their reciprocal behavior towards the trust received from first movers is not associated with their other-regarding preferences. Trustees' attitudes towards inequality, elicited in the context of both advantageous and disadvantageous inequality tests, has no effect on their decision to reciprocate towards trusting first movers. One possible explanation may be found in the differences between the two dictator game roles used to elicit subjects' inequality aversion in the AIA and DIA tasks, and the single-role played in the TG. Whereas in the elicitation inequality tasks individuals play both roles, the dictator and the recipient, being one of the roles randomly chosen for the final payoff, in the TG subjects play only one role, trustor or trustee, exogenously determined. To the extent that second movers do not hold themselves responsible for the distribution of roles, they could have avoided ex-ante fairness criteria. This argument is defended by the accountability principle in the theory of justice, which states that subjects may not feel responsible by factors beyond individual control (Cappelen *et al*., 2007). Instead, some personality traits are found to be significantly associated with reciprocal behavior. Specifically, capturing the propensity to cooperate in conflictive situations,

higher agreeableness predicts a higher probability of reciprocal behavior by second movers. However, the hypothesis concerning the relevance of agreeableness on trust by means of trustworthiness expectations via social projection, is not confirmed. Our results corroborate the arguments used by Krueger (2013) in the sense of rejecting the validity of this theory in social dilemmas in which players move sequentially or payoffs are asymmetrical. In such a type of games the first mover can ask what she would do in the second mover's position and project that inclination. In the TG however, it is difficult to project own preferences onto the trustee's reciprocation since trustor and trustee roles are essentially different. Furthermore, it is a robust finding in the literature that subjects strongly project their own behaviors onto in-group members, but hardly onto outgroup members (Krueger *et al*., 2008).

Acknowledgments

Financial support by Universitat Jaume I (grant UJI-B2018-76) and by the Spanish *Ministerio de Economía y Competitividad* (grant ECO2015-68469-R) and *Ministerio de Ciencia, Innovación y Universidades* (grant RTI2018-096927-B-I00) is gratefully acknowledged.

References

Aimone, J. A., Houser, D., & Weber, B. (2014). Neural signatures of betrayal aversion: An fMRI study of trust. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1782). https://doi.org/10.1098/rspb.2013.2127

Alarcon, G. M., Lyons, J. B., Christensen, J. C., Bowers, M. A., Klosterman, S. L., & Capiola, A. (2018). The role of propensity to trust and the five factor model across the trust process. *Journal of Research in Personality*, *75*, 69–82. https://doi.org/10.1016/j.jrp.2018.05.006

Anderson, L. R., Mellor, J. M., & Milyo, J. (2006). Induced heterogeneity in trust experiments. *Experimental Economics*, *9*, 223–235. https://doi.org/10.1007/s10683-006-9124-2

Ashraf, N., Bohnet, I., & Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, *9*(3), 193–208. https://doi.org/10.1007/s10683-006-9122-4

Attanasi, G., Georgantzís, N., Rotondi, V., & Vigani, D. (2018). Lottery- and survey-based risk attitudes linked through a multichoice elicitation task. *Theory and Decision*, *84*(3), 341–372. https://doi.org/10.1007/s11238-017-9613-0

Barreda-Tarrazona, I., Sabater-Grande, G., & Georgantzís, N. (2020). *Risk elicitation through the S-GG lottery panel task: Implementation note.* Working Papers 2020/23, Economics Department, Universitat Jaume I, Castellón (Spain).

Beattie, T. (2019). *Decomposing trust into risk preferences, altruism, and subjective beliefs: An experimental analysis*. Master Dissertation. Faculty of Commerce, School of Economics, University of Cape Town.

Ben-Ner, A., & Halldorsson, F. (2010). Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology*, *31*(1), 64–79. https://doi.org/10.1016/j.joep.2009.10.001

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142. https://doi.org/10.1006/game.1995.1027

Blanco, M., Engelmann, D., & Normann, H. T. (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior*, *72*(2), 321–338. https://doi.org/10.1016/j.geb.2010.09.008

Bohnet, I., & Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior and Organization*, *55*(4 SPEC.ISS.), 467–484. https://doi.org/10.1016/j.jebo.2003.11.004

Brañas-Garza, P., Jiménez-Jiménez, F., & Morales, A. J. (2004). Strategic uncertainty and risk attitudes: The experimental connection, *Cuadernos de Economía, 27*, 139-152.

Brülhart, M., & Usunier, J. C. (2012). Does the trust game measure trust? *Economics Letters*, *115*(1), 20–23. https://doi.org/10.1016/j.econlet.2011.11.039

Burks, S. V., Carpenter, J. P., & Verhoogen, E. (2003). Playing both roles in the trust game. *Journal of Economic Behavior and Organization*, *51*(2), 195–216. https://doi.org/10.1016/S0167-2681(02)00093-8

Cappelen, A. W., Hole, A. D., Sørensen, E. Ø., & Tungodden, B. (2007). The Pluralism of Fairness Ideals: An Experimental Approach. *American Economic Review*, *97*(3), 818–827. https://doi.org/10.1257/AER.97.3.818

Chetty, R., Hofmeyr, A., Kincaid, H., & Monroe, B. (2021). The Trust Game does not (only) measure trust: The risk-trust confound revisited. *Journal of Behavioral and Experimental Economics*, *90*, 101520. https://doi.org/10.1016/j.socec.2020.101520

Costa, P. T., & McCrae, R. R. (2008). The revised NEO personality inventory (NEO-PI-R). In *The SAGE Handbook of Personality Theory and Assessment: Volume 2 - Personality Measurement and Testing* (pp. 179–198). https://doi.org/10.4135/9781849200479.n9

Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*, *46*(2), 260–281. https://doi.org/10.1016/S0899-8256(03)00119-2

Denissen, J. J. A., & Penke, L. (2008). Motivational individual reaction norms underlying the Five-Factor model of personality: First steps towards a theory-based conceptual framework. *Journal of Research in Personality*, *42*(5), 1285–1302. https://doi.org/10.1016/j.jrp.2008.04.002

Eckel, C. C., & Wilson, R. K. (2004). Is trust a risky decision? *Journal of Economic Behavior and Organization*, *55*(4), 447–465. https://doi.org/10.1016/j.jebo.2003.11.003

Fahr, R., & Irlenbusch, B. (2008). Identifying personality traits to enhance trust between organisations: An experimental approach. *Managerial and Decision Economics*, *29*(6), 469–487. https://doi.org/10.1002/mde.1415

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868. https://doi.org/10.1162/003355399556151

Fetchenhauer, D., & Dunning, D. (2012). Betrayal aversion versus principled trustfulness—How to explain risk avoidance and risky choices in trust games. *Journal of Economic Behavior and Organization*, *81*(2), 534–541. https://doi.org/10.1016/j.jebo.2011.07.017

Gambetta, D. (1988) Can we trust trust? In: Gambetta, D., Ed., *Trust: Making and Breaking Cooperative Relations*. Blackwell, New York, 213-237.

Garapin, A., Muller, L., & Rahali, B. (2015). Does trust mean giving and not risking? Experimental evidence from the trust game. *Revue d'Economie Politique*, *125*(5), 701–716. https://doi.org/10.3917/redp.255.0701

Griesinger, D. W., & Livingston, J. W. (1973). Toward a model of interpersonal motivation in experimental games. *Behavioral Science*, *18*(3), 173–188. https://doi.org/10.1002/bs.3830180305

Gunnthorsdottir, A., McCabe, K., & Smith, V. (2002). Using the Machiavellianism instrument to predict trustworthiness in a bargaining game. *Journal of Economic Psychology*, *23*(1), 49–66. https://doi.org/10.1016/S0167-4870(01)00067-8

Hargreaves Heap, S. P., Tan, J. H. W., & Zizzo, D. J. (2013). Trust, inequality and the market. *Theory and Decision*, *74*(3), 311–333. https://doi.org/10.1007/s11238-011-9287-y

Houser, D., Schunk, D., & Winter, J. (2010). Distinguishing trust from risk: An anatomy of the investment game. *Journal of Economic Behavior and Organization*, *74*(1–2), 72–81. https://doi.org/10.1016/j.jebo.2010.01.002

Ibáñez, M. I., Sabater-Grande, G., Barreda-Tarrazona, I., Mezquita, L., López-Ovejero, S., Villa, H., Perakakis, P., Ortet, G., García-Gallego, A., & Georgantzís, N. (2016). Take the money and run: Psychopathic behavior in the trust game. *Frontiers in Psychology*, *7*, 1866. https://doi.org/10.3389/fpsyg.2016.01866

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five Trait Taxonomy history, measurement, and conceptual issues. In *Handbook of Personality: Theory and Research* (pp. 114–158). New York: Guilford Press.

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *32*(5), 865–889. https://doi.org/10.1016/j.joep.2011.05.007

Krueger, J. I., Massey, A. L., & DiDonato, T. E. (2008). A matter of trust: From social preferences to the strategic adherence to social norms. *Negotiation and Conflict Management Research*, *1*(1), 31–52. https://doi.org/10.1111/j.1750-4716.2007.00003.x

Krueger, J. I. (2013). Social projection as a source of cooperation. *Current Directions in Psychological Science*, *22*, 289-294. https://doi.org/10.1177/0963721413481352

Lee, W.-S., & Selart, M. (2015). How betrayal affects emotions and subsequent trust. *The Open Psychology Journal*, *8*, 153–159.

Lei, V., & Vesely, F. (2010). In-group versus out-group Trust: The impact of income inequality. *Southern Economic Journal*, *76*(4), 1049–1063. https://doi.org/10.4284/sej.2010.76.4.1049

McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *The Academy of Management Review*, *23*(3), 473. https://doi.org/10.2307/259290

Müller, J., & Schwieren, C. (2020). Big Five personality factors in the trust game. *Journal of Business Economics*, *90*(1), 37–55. https://doi.org/10.1007/s11573-019-00928-3

Nguyen, Q., Villeval, M. C., & Xu, H. (2016). Trust under the prospect theory and quasi-hyperbolic preferences: A field experiment in Vietnam. *Economic Development and Cultural Change*, *64*(3), 546–572. https://doi.org/10.1086/685434

Özer, Ö., & Zheng, Y. (2017). Trust and Trustworthiness. In K. Donohue, E. Katok, & S. Leider (Eds.), *The Handbook of Behavioral Operations*. Wiley. https://doi.org/10.2139/ssrn.3046303

Rodriguez-Lara, I. (2018). No evidence of inequality aversion in the investment game. *PLOS ONE*, *13*(10), e0204392. https://doi.org/10.1371/journal.pone.0204392

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393–404. https://doi.org/10.5465/AMR.1998.926617

Sabater-Grande, G., & Georgantzis, N. (2002). Accounting for risk aversion in repeated prisoners' dilemma games: An experimental test. *Journal of Economic Behavior and Organization*, *48*(1), 37–50. https://doi.org/10.1016/S0167-2681(01)00223-2

Sabater-Grande, G., Haro, G., García-Gallego, A., Georgantzís, N., Herranz-Zarzoso, N., & Baquero, A. (2020). Risk-taking and fairness among cocaine-dependent patients in dual diagnoses: Schizophrenia and Anti-Social Personality Disorder. *Scientific Reports*, *10*(1), 1–13. https://doi.org/10.1038/s41598-020-66954-2

Sapienza, P., Toldra-Simats, A., & Zingales, L. (2013). Understanding trust. *The Economic Journal*, *123*(573), 1313–1332. https://doi.org/10.1111/ecoj.12036

Schechter, L. (2007). Traditional trust measurement and the risk confound: An experiment in rural Paraguay. *Journal of Economic Behavior and Organization*, *62*(2), 272–292. https://doi.org/10.1016/j.jebo.2005.03.006

Swope, K. J., Cadigan, J., Schmitt, P. M., & Shupp, R. (2008). Personality preferences in laboratory economics experiments. *Journal of Socioeconomics*, *37*(3), 998–1009. https://doi.org/10.1016/j.socec.2006.12.065

Szijjártó, L., Kocsor, F., & Bereczkei, T. (2018). Machiavellian individuals' reciprocation tends to be smaller in a Trust Game. *Human Ethology Bulletin*, *33*, 39–48. https://doi.org/10.22330/heb/331/039-048

Thielmann, I., & Hilbig, B. E. (2014). Trust in me, trust in you: A social projection account of the link between personality, cooperativeness, and trustworthiness expectations. *Journal of Research in Personality*, *50*(1), 61–65. https://doi.org/10.1016/j.jrp.2014.03.006

Thielmann, I., & Hilbig, B. E. (2015a). Trust: An integrative review from a person-situation perspective. *Review of General Psychology*, *19*(3), 249–277. https://doi.org/10.1037/gpr0000046

Thielmann, I., & Hilbig, B. E. (2015b). The traits one can trust: Dissecting reciprocity and kindness as determinants of trustworthy behavior. *Personality and Social Psychology Bulletin*, *41*(11), 1523–1536. https://doi.org/10.1177/0146167215600530

Xiao, E., & Bicchieri, C. (2010). When equality trumps reciprocity. *Journal of Economic Psychology*, *31*(3), 456–470. https://doi.org/10.1016/j.joep.2010.02.001

TABLES

Table 1. SGG Lottery-Panel Task

Panel 1										
Prob.	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
€	1.00	1.10	1.30	1.50	1.70	2.10	2.70	3.60	5.40	10.90
Panel 2										
Prob.	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
€	1.00	1.20	1.50	1.90	2.30	3.00	4.00	5.70	9.00	19.00
Panel 3										
Prob.	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
€	1.00	1.70	5.50	3.60	5.00	7.00	10.00	15.00	25.00	55.00
Panel 4										
Prob.	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
€	1.00	2.20	3.80	5.70	8.30	12.00	17.50	26.70	45.00	100.00
Panel 5										
Prob.	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
€	0.00	0.10	0.30	0.50	0.70	1.10	1.70	2.60	4.40	9.90
Panel 6										
Prob.	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
€	0.00	0.20	0.50	0.90	1.30	2.00	3.00	4.70	8.00	18.00
Panel 7										
Prob.	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
€	0.00	0.70	1.50	2.60	4.00	6.00	9.00	14.00	24.00	54.00
Panel 8										
Prob.	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
€	0.00	1.20	2.80	4.70	7.30	11.00	16.50	25.70	44.00	99.00

Table 2. Advantageous Inequality Aversion (AIA) task

Row	My gains	Partner's gains	My gains	Partner's gains
1	€15	€15	€15	€0
2	€14	€14	€15	€0
3	€13	€13	€15	€0
4	€12	€12	€15	€0
5	€11	€11	€15	€0
6	€10	€10	€15	€0
7	€9	€9	€15	€0
8	€8	€8	€15	€0
9	€7	€7	€15	€0
10	€6	€6	€15	€0
11	€5	€5	€15	€0
12	€4	€4	€15	€0
13	€3	€3	€15	€0
14	€2	€2	€15	€0
15	€1	€1	€15	€0
16	€0	€0	€15	€0

Table 3. Disadvantageous Inequality Aversion (DIA) task

Row	My gains	Partner's gains	My gains	Partner's gains
1	€15	€15	€0	€15
2	€14	€14	€0	€15
3	€13	€13	€0	€15
4	€12	€12	€0	€15
5	€11	€11	€0	€15
6	€10	€10	€0	€15
7	€9	€9	€0	€15
8	€8	€8	€0	€15
9	€7	€7	€0	€15
10	€6	€6	€0	€15
11	€5	€5	€0	€15
12	€4	€4	€0	€15
13	€3	€3	€0	€15
14	€2	€2	€0	€15
15	€1	€1	€0	€15
16	€0	€0	€0	€15

Table 4: Descriptive statistics of the explanatory variables considered

N=220	SGG P1-P4	SGG P5-P8	AIA	DIA	A	N
Mean	0.44	0.49	10.30	4.69	116.50	92.92
S.D.	0.17	0.21	3.78	4.92	18.61	23.40
S-W test	0.065*	0.014**	0.000***	0.000***	0.014**	0.375

Note: SGG P1-P4: per-subject mean winning probability of the lotteries chosen in panels 1 to 4; SGG P5-P8: per-subject mean winning probability of the lotteries chosen in panels 5 to 8; AIA: advantageous inequity aversion; DIA: disadvantageous inequity aversion; A: agreeableness; N: neuroticism. *p < 0.10; **p< 0.05; ***p < 0.01.

Table 5. Probit estimation for Trustors

Variable	**Trust**
SGG P1-P4	-0.5053*
	(0.299)
SGG P5-P8	0.3314
	(0.241)
AIA	0.0015
	(0.0142)
DIA	0.0041
	(0.0101)
A	0.0024
	(0.0027)
N	-0.0028
	(0.0023)
Gender	0.0792
	(0.1171)
Observations	110

Note: SGG P1-P4: per-subject mean winning probability of the lotteries chosen in panels 1 to 4; SGG P5-P8: per-subject mean winning probability of the lotteries chosen in panels 5 to 8; AIA: advantageous inequality aversion; DIA: disadvantageous inequality aversion; A: agreeableness; N: neuroticism. Coefficients correspond to marginal effects. Standard errors in parentheses; * $p<0.10$.

Table 6. Probit estimation for Trustees

Variable	**Reciprocation**
AIA	0.0039
	(0.0183)
DIA	-0.0033
	(0.0141)
A	0.0077*
	(0.0044)
Gender	-0.0372
	(0.1661)
Observations	52

Note: AIA: advantageous inequality aversion; DIA: disadvantageous inequality aversion; A: agreeableness. Standard errors in parentheses; * p<0.10.