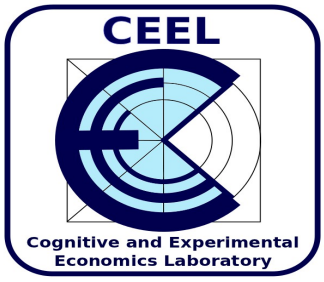


Riccardo Ghidoni and Matteo Ploner

When do the Expectations of Others Matter? An Experiment on Distributional Justice and Guilt Aversion

CEEL Working Paper 3-14

Cognitive and Experimental
Economics Laboratory

Via Inama, 5 38122 Trento, Italy

http://www-ceel.economia.unitn.it
tel. +39.461.282313

When do the Expectations of Others Matter?

An Experiment on Distributional Justice and Guilt Aversion*

Riccardo Ghidoni† Matteo Ploner‡

Abstract

In a modified dictator game experiment, we study how distributional justice, measured by the proportionality between effort exerted and rewards obtained, and guilt feelings triggered by others' expectations affect dictator's choices. We consider these two sources of behavior in isolation and in interaction. Our results suggest that both justice concerns and guilt aversion are important drivers of behavior. However, the expectations of others are more relevant when the choice environment is likely to induce less equitable outcomes.

Keywords: Justice; Guilt Aversion; Entitlement Rights; Experiments.

*We would like to thank Martin Dufwenberg, Marco Casari, Maria Bigoni, Stefania Bortolotti, Ennio Bilancini, Valeria Maggian, and Alessandro Saia for their useful comments. The paper also benefited from the comments of participants at the VII G.R.A.S.S. Workshop, the Trento Workshop on Social Norms and Social Preferences, and seminars at the University of Bologna and at the University of Innsbruck. The usual caveat applies. Our thanks also go to EIEF for the financial support.

†Department of Economics, University of Bologna. Email: `riccardo.ghidoni2@unibo.it`. Web: `https://sites.google.com/site/riccardoghidoni/`

‡Department of Economics and Management, University of Trento. Email: `matteo.ploner@unitn.it`. Web: `https://sites.google.com/site/mploner/`

1 Introduction

In a well-known passage of his Second Treatise on Government, John Locke writes: "I think, it is very easy to conceive, without any difficulty, how labour could at first begin a title of property in the common things of nature [...]" (Chapter 5, Section 51). This sentence captures the essence of what Locke labels a law of nature, i.e. that property rights on goods originate directly from effort exerted to generate them. The Lockean law of nature is grounded on a basic justice principle according to which outcomes should be related to actions (desert). This general distributional principle has been subject to extensive philosophical debate and has attracted the attention of experimental research, both in social psychology (e.g. Leventhal and Michaels, 1969) and in economics (e.g. Hoffman and Spitzer, 1985; Konow, 2000).

We aim at assessing the relevance of the "hard-wired" justice principle that relates actions and consequences against an alternative measure of justice driven by emotions originating in social interactions. Emotions have indeed been recognized to have a major influence on economic behavior (Elster, 1998). In the last decade, the emotion of guilt has received attention both from a theoretical and experimental perspective, due also to the theory of *guilt aversion* (e.g. Dufwenberg and Gneezy, 2000; Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007; Ellingsen et al., 2010).[1] According to this theory, decision-makers can experience a negative feeling, i.e. guilt, whenever they believe that their action will contribute to let their counterpart down.

Relying on a within-subjects experimental design, we endeavor to understand if and how justice concern à la Locke and guilt aversion interact in shaping the behavior of decision-makers. Our main objective is to test whether expectations of the counterpart about the behavior of the decision-makers affect decision-makers' behavior only when these expectations are not in conflict with justice considerations. In other words, justice considerations may be key to understanding when others' expectation are seen as *legitimate* by the decision-maker and, thus, worth taking into account.

Our data show that both guilt feelings and justice considerations play a fundamental role in explaining the choices of decision-makers. However, quite unexpectedly, others'

[1] Although the focus of the present paper is on the theory of guilt aversion as formalized by Charness and Dufwenberg (2006) and Battigalli and Dufwenberg (2007), we are aware that the emotion of guilt has received attention in other settings (e.g. Lindbeck and Nyberg, 2006; Cervellati et al., 2010).

expectations become more relevant when the choice environment offers less protection to entitlement rights originated by effort and, thus, it is less likely to produce equitable outcomes. In the concluding section, we provide a possible explanation for this result which contradicts our main research hypothesis. We also draw attention to how further research on the interaction between institutional protection of entitlement rights and subjective expectations about standards of behavior is needed.

The rest of the paper is organized as follows: in Section 2, we present a brief literature review of desert and of guilt aversion theories; in Section 3, we present our experimental game and our research hypotheses and in Section 4 we describe the experiment; in Section 5 we illustrate the results from our experiment and in Section 6 we discuss these results and provide some final remarks.

2 Related Literature

2.1 Distributional Justice

Konow (2003) presents an extensive review of justice concepts that are relevant to economics and provides the empirical basis to build a descriptive theory of justice. The author highlights the relevance of theories that appeal to desert and relate fair allocations to individual actions. Within this class of models, equity theory (e.g. Adams, 1963) provides guidance to assess the fairness of allocations in which a production stage is involved. The basic tenet of the theory is that an equitable allocation should preserve the proportionality of resources invested (input) and rewards obtained (output) across individuals. Thus, those investing more resources in the production of the output should obtain more out of it than those investing less. Hoffman and Spitzer (1985) point out how the proportionality principle underlying equity theory captures the essence of the Lockean natural law. Empirical support for the relevance of this justice principle is provided by experimental studies in social psychology (e.g. Leventhal and Michaels, 1969; Mikula, 1974) and economics (e.g. Konow, 2000).

To establish whether an allocation is fair, it is crucial to define the nature of the input against which the output is evaluated. According to attribution theory (e.g. Weiner, 1985), only factors that are directly controlled by individuals qualify to establish the fairness of an allocation. This principle of justice is captured also by the accountability principle of Konow (1996), which distinguishes between discretionary variables and ex-

ogenous variables. According to Konow, only the former should be taken into account when assessing the fairness of allocations. The recent experimental work by Becker (2013) shows that normative beliefs are strongly influenced by accountability considerations, but individuals often violate the principle and selfishly enjoy rewards that originate from pure luck. These results are also supported in the structural estimation presented by Cappelen et al. (2007), when comparing alternative fairness ideals. The authors identify a plurality of fairness ideals in the population, but also show that a non negligible share of the population maintains a libertarian view according to which individuals deserve what they produce, irrespective of the control exerted over the production factors.

2.2 Guilt Aversion

The idea of guilt aversion understood as the aversion to letting others down was first introduced by Dufwenberg and Gneezy (2000), who observe, both in a lost wallet game and in a dictator game, a positive correlation between trustees' (dictators') transfers and their second-order beliefs elicited after the play.[2] In a later experiment, also Bacharach et al. (2007) find evidence supporting this definition of guilt aversion in three modified trust games. Charness and Dufwenberg (2006) and Battigalli and Dufwenberg (2007) develop a formal model of guilt aversion based on the analytical framework of the *psychological games* (Geanakoplos et al., 1989; Battigalli and Dufwenberg, 2009).[3]

An important application of guilt aversion is provided by Charness and Dufwenberg (2006), who exploit this theory to explain why individuals in experiments tend to keep non-binding promises even when they would earn more by not doing so. Specifically, Charness and Dufwenberg propose a modified trust game that allows for moral hazard by trustees (so-called trust game *with hidden action*), where subjects can also exchange free form messages before the play. Charness and Dufwenberg first observe a positive correlation between promises and trustworthiness. Moreover, consistently with guilt aversion, they find that trustworthy choices are positively correlated with trustees' second-order beliefs elicited after the play. Thus, Charness and Dufwenberg conclude that promises

[2]The lost wallet game is a modified trust game which allows for several wealth multipliers, and where the trustor only faces a dichotomous choice between trusting the trustee or not.

[3]Battigalli and Dufwenberg (2007) introduce a distinction between *simple guilt* and *guilt from blame*, where the main difference between the two is in the fact that in the second form of guilt aversion, the decision-maker dislikes being blamed, rather than simply letting the counterpart down.

foster trustees' second-order beliefs about trustors' first-order belief on trustworthiness, triggering a sense of guilt in those trustees who let trustors down.

Charness and Dufwenberg's conclusions are questioned by Vanberg (2008), who tests whether individuals keep their promises because they dislike letting others down, as predicted by guilt aversion, or rather because they have a taste for keeping their own word (so-called "commitment-based explanation"), as modeled by Ellingsen and Johannesson (2004) and Kartik (2009). Using a design that closely resembles that of Charness and Dufwenberg's, Vanberg finds support for the commitment-base explanation. In a trust game with hidden-action, where subjects could only send bare promises, Charness and Dufwenberg (2010) test whether "truth-value [is] all we need to capture an important aspect of human motivation, or does the context in which the statement was made matter". These authors find limited support both for a commitment-based explanation and for guilt aversion.

Ellingsen et al. (2010) challenge Charness and Dufwenberg's conclusions from another perspective arguing that the positive correlation between trustees' second-order beliefs and their back-transfers is mainly due to the so-called "false consensus" effect (Engelmann and Strobel, 2000, 2011), rather than to guilt aversion.[4] According to this alternative explanation, trustees may consider their own back-transfer choice as the most representative, and hence believe that trustors' first-order beliefs will coincide with their own. Thus, the same trustees who prefer to make a large back-transfer could be those who believe that trustors expect a large back-transfer. To test this hypothesis, instead of eliciting second-order beliefs after the play, Ellingsen et al. reveal their partner's first-order belief to those subjects who were supposed to suffer from guilt. Ellingsen et al. do not detect any evidence of guilt aversion in any of their experimental games, i.e. dictator game, lost-wallet game, and trust game. Reuben et al. (2009), however, proposes a cleaner version of the design by Ellingsen et al. and find that trustees are less trustworthy when trustors have low expectations, concluding that mis-trust is self-fulfilling.[5] Additional evidence in

[4]Vanberg also remarks that the correlation between second-order beliefs and actions could be driven by the false consensus effects.

[5]Reuben et al. operate three major changes on the Ellingsen et al.'s design: (1) they provide all subjects with the same instructions; (2) all subjects play both as trustor and trustee sequentially, and beliefs are elicited when subjects play the role of senders, in a round when beliefs are unused; (3) they increase the reward for the accuracy of beliefs and the game payoffs.

support of guilt aversion in a different experimental setting, is offered by Ockenfels and Werner (2014) who study how different scales with which beliefs are elicited before the game can subsequently affect subjects' behavior. Their findings also seem to suggest that, in dictator games, dictators are motivated by not letting recipients' down rather than by conforming to a norm regarding dictators' behavior.

Recently, Battigalli et al. (2013) proved how guilt aversion can offer an explanation for the findings of the seminal experiment on deception by Gneezy (2005). Dufwenberg et al. (2011) show how framing effect can influence contributions in a framed public goods game through changes in subjects' second-order beliefs. Finally, Bellemare et al. (2011) estimate structural models of guilt from a large scale sequential game, where subjects are willing to pay between forty to eighty cents to avoid letting their partner down by one euro.

3 Entitlement, Justice, and Guilt

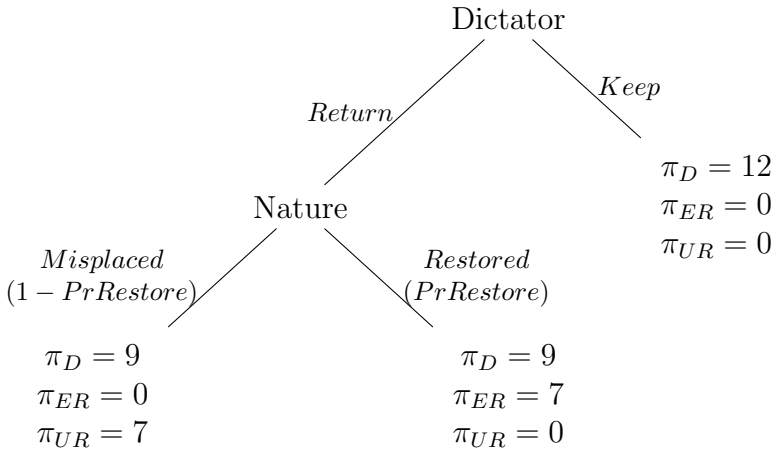
3.1 A Modified Dictator Game

In our modified dictator game (Figure 1), one player (i.e. the dictator) who exerted an effort to generate her endowment, or "wallet", can choose whether to *Return* or to *Keep* the wallet generated by another player (i.e. the entitled recipient), who randomly lost it, thus remaining with no reward for the work done. When making a choice, each dictator faces a certain probability $PrRestore$ with which the entitled recipient restores her wallet, conditional upon the dictator choosing to return it. With alternative "restoring probabilities" 4/6, 5/6, or 6/6, the wallet is returned to the entitled recipient. Otherwise, the wallet is misplaced by Nature to a third dummy player who exerted no effort (i.e. the unentitled recipient) with probabilities of 2/6, 1/6, or 0/6, respectively.[6] Only the dictator knows the specific restoring probability she is facing in a certain round, while recipients do not know. Moreover, the entitled recipients do not observe the action of the dictator, so she cannot infer whether she did not received her wallet back because the

[6]The payoffs in the game ensure that under alternative restoring probabilities neither efficiency nor surplus distribution change. Thus, changes in restoring probabilities should not affect the behavior of dictators characterized by efficiency and/or fairness concerns. Furthermore, notice that choosing *Keep* is inefficient since the payoff obtained by the dictator in this case is lower than the sum of two wallets, i.e. 12 ECU instead of 14 ECU.

dictator chose to keep it or because the Nature misplaced it.[7] In this context, Nature hence captures (exogenous) institutional aspects, known only to the decision-maker, which can characterize the choice environment by ensuring different degrees of protection of the entitlement rights.

Figure 1: Modified Dictator Game.



Note: D: dictator; ER: entitled recipient; UR: unentitled recipient.

When choosing whether or not to return the wallet, the dictator is not only informed about the restoring probability associated to each of her choices, but also about the overall expectations of the entitled recipient regarding the return of the wallet. Recipients are asked how many times, out of the three decisions made by a dictator, the wallet will be returned. The expectations of the recipients, then reported to the dictator, can go from very optimistic, i.e. believing that the wallet is always returned, to very pessimistic, i.e. believing that the wallet is never returned (more details about the beliefs elicitation procedure are given in Section 4).

3.2 Research Hypotheses

Under standard assumptions of selfish rationality, dictators in our game should never choose to return the wallet to an entitled recipient, irrespective of both the probability with which the wallet gets misplaced and the entitled recipient's expectation. However, a large body of literature has highlighted how social preferences may play an important role

[7]The presence of the Nature and the unobservability of the dictator's action are both features that are also present in the games of Charness and Dufwenberg (2006) and Vanberg (2008).

in determining the behavior of individuals participating in experiments. In our context, outcome-based social preferences like altruism (e.g. Cox et al., 2008), inequity-aversion (e.g. Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999), or efficiency (e.g. Charness and Rabin, 2002) may induce the dictator to return the wallet. Thus, unlike under assumptions of selfish rationality, a large class of social preferences predicts that dictators may choose to return the wallet. However, like under selfish rationality, outcome-based social preferences predict that dictators decisions should be neither affected by different restoring probabilities nor by the expectations of the entitled recipient.

While selfishness and outcome-based social preferences provide us with useful benchmark behavioral predictions, the main focus of our inquiry is on how justice concerns linked to the protection of entitlement rights (proxied by restoring probabilities) and counterparts' expectations (measured by recipients' first-order beliefs) affect dictators' decisions. Accordingly, we present some testable hypotheses which refer to the impact of these potential sources of behavior.

When a dictator "finds" the wallet lost by an entitled recipient, a strongly unfair and inequitable allocation is exogenously induced because both subjects have exerted the same effort, but (almost) all the surplus generated is given to the dictator. Dictators can restore justice by returning the wallet for which the entitled recipient worked. However, when the returned wallet is misplaced because of an interference of Nature, an even less equitable allocation is in place as the wallet is given to someone who did not exert any effort at all. Therefore, alternative restoring probabilities impact on the anticipated proportionality between efforts exerted to "generate" the wallet and rewards obtained when the wallet is returned.

According to equity theory (Adams, 1963) and to the accountability principle (Konow, 2000, 2003), individuals investing more resources in the production of a certain surplus should receive a larger share of it than those investing less. In our context, justice concerns based on input/output proportionality translate into a higher likelihood of returning the wallet when the chances of a misplaced return are lower. Thus, relying on this well-established principle of equity, we formulate the following hypotheses.

Hypothesis 1. ***Justice Concerns***

The higher the restoring probability (i.e. the lower the likelihood of a misplaced return), the more likely dictators are to choose to return the wallet.

According to the theory of guilt aversion (Charness and Dufwenberg, 2006; Battigalli

and Dufwenberg, 2007), an individual will be disappointed when the final outcome of the game does not match expectations. Thus, decision-makers who can influence the outcome of the game may experience a negative emotion of guilt whenever their action contributes to letting their counterpart down. A guilt sensitive decision-maker may hence prefer to renounce part of her material payoff to avoid the psychological cost of guilt. Given that in our experiment recipients are kept blind about the restoring probability faced by the dictator, more optimistic recipients will always be more disappointed by realizing that they did not receive their wallet back. The feeling of guilt experienced when retaining the wallet should hence be stronger when facing an optimistic recipient than when facing a less optimistic one, irrespective of the restoring probability. Since in our context dictators freely access the expectations of the entitled recipient, the theory of guilt aversion translates into the following testable hypothesis.[8]

Hypothesis 2. ***Guilt Aversion***

The more optimistic the recipients' expectations about the return of the wallet, the more likely dictators are to return it.

Hypothesis 1 and 2 provide us with a guidance to evaluate the impact of equity considerations and of recipient's expectation in isolation. However, these two sources of behavior are likely to interact in shaping dictator's behavior. Even though guilt aversion predicts dictators to suffer the same amount of guilt irrespective to the restoring probability, we expect dictators to attach a different weight to the same expectation when facing different restoring probabilities. The intuition here is that the impact of counterpart's expectation on dictator's choices is conditional upon environmental features, which may remain unknown to the former but not to the latter. To elaborate, beliefs of the counterpart may be legitimate in the eyes of the dictator only when these beliefs are not in conflict with justice considerations. As pointed out in Hypothesis 1, the share of dictators ready to return the wallet should be lower when a misplaced return is more likely. Thus, returning the wallet when an unfair outcome is more likely should be considered less appropriate

[8]Similar to Charness and Dufwenberg (2006), Hypothesis 2 in our paper presumes that "players do not coordinate on some "equilibrium"; it refers only to the individual player and properties of his/her utility." In particular, we assume that recipients do not perfectly anticipate the degree of guilt sensitivity of the dictators they meet across the three rounds. This assumption seems reasonable since guilt sensitivity can differ across individuals (Tangney, 1995). Moreover, we *never* provide recipients with feedback about dictators' choices in order to avoid any belief update.

than when the unfair outcome is less likely. Accordingly, we expect that higher chances of a misplaced return will reduce the impact of the entitled recipient's expectations on the decision to return because the sense of guilt originated by the decision to keep the wallet is conditional upon justice considerations.[9] In terms of a testable hypothesis, we expect to register a positive interaction between the counterpart's expectations and restoring probabilities in affecting returning choices.

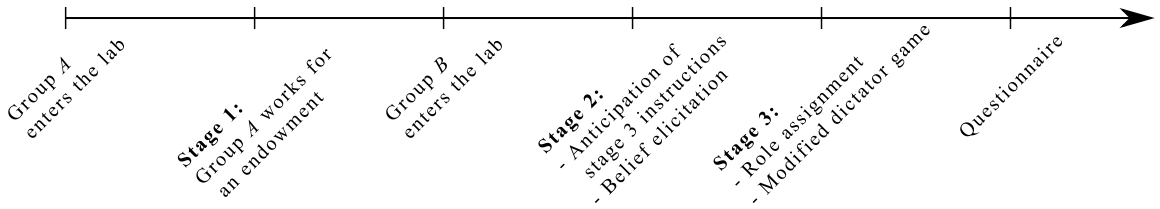
Hypothesis 3. ***Conditional Guilt Aversion***

The positive impact of recipients' expectations on the choice to return the wallet is stronger when the restoring probability is higher (i.e. when a misplaced return is less likely).

4 Experimental Design

Two separate groups of subjects take part in each experimental session, group A (10 subjects) and group B (5 subjects). While members of group A actively participate in all the three stages of the experiment, group B members actively participate only in the second stage (i.e. beliefs elicitation). Members of group B enter the laboratory after Stage 1 and are allowed to surf the Internet during Stage 3. Figure 2 summarizes the timeline of the experimental session.

Figure 2: Timeline of the Experimental Session.



In Stage 1, all members of group A perform a task to "generate" their own endowment. The task consist in counting the number of zeros in seven 15×8 tables that sequentially appear on computer screens, and contain both 0 and 1 digits in random proportions.[10] For

[9] In principle, also the reverse may be true, i.e. justice considerations are conditional upon the negative emotion of guilt experienced when taking a decision. Our design does not allow us to shed light on the direction in which these two behavior sources interact in shaping the dictators' behavior.

[10] We borrowed this task from Abeler et al. (2011) to induce a sense of entitlement on wealth in the experiment. Previous works have shown the relevance of asset legitimacy in simple bargaining situations.

each table solved, the subject earns 1 ECU (Experimental Currency Unit, 1 ECU=€1). Subjects are not time constrained, and are allowed to make mistakes and retry to enter the correct number of zeros. Thus, at the end of the first stage, each member of group A virtually owns an endowment of 7 ECUs.[11]

After group B has also entered the laboratory, the experimental game played in Stage 3, which consists in the modified dictator game depicted in Figure 1, is carefully explained to all participants. In the game, subjects are matched in triplets, comprising two members of group A and one member of group B. We exploit the following mechanism to randomly assign subjects to their role. Before playing, half of group A members are randomly chosen and "lose" their wallet. These subjects become entitled recipients. Lost wallets are "found" by the other half of group A members, who become dictators. The third member of a triplet, i.e. the group B member, is the unentitled recipient, who is allowed to surf the Internet during Stage 3.

The game is repeated three times, during which subjects are matched with a *perfect stranger protocol*. The order in which the restoring probabilities are faced by dictators is random, and it remains unknown to recipients. No feedback is given either to dictators or to recipients at the end of each round. At the end of the session, one of the three rounds is randomly drawn for the payment. Dictators are informed about the final outcome of the round drawn for the payment, while recipients only learn the final amount of ECUs they own. Thus, entitled recipients will not be able to distinguish if the dictator did not choose to return the wallet, or if the wallet was misplaced in the hands of the unentitled recipient.

Since we are interested in how guilt aversion may affect dictators' behavior, belief elicitation represents a crucial stage of the design. Unlike other experiments (e.g. Dufwenberg and Gneezy, 2000; Charness and Dufwenberg, 2006), we do not test for dictators' guilt aversion by eliciting second-order beliefs after the play to avoid identification problems due to false consensus effects (see Engelmann and Strobel, 2000, 2011; Ellingsen et al.,

As an example, Hoffman et al. (1994) and Cherry et al. (2002) administered a knowledge questionnaire to participants before taking part in a ultimatum and dictator game respectively, and show that this strongly reduces other-regarding concerns. We chose the Abeler et al. task because it provides a more direct measure of effort exerted relative to a knowledge test in which human capital and luck seem to be more important than effort.

[11] In the instructions, we do not use the word "wallet" but we stress that each subject earns her *own* endowment by performing the task.

2010, for a discussion of its importance when testing guilt aversion). We prefer to induce guilt feelings by providing dictators with entitled recipient's first-order beliefs.[12] To this end, in Stage 2 (i.e. previous to role assignment), we ask *all* subjects to anticipate how many times out of the three rounds they expect a *generic* dictator to return the wallet.

Table 1 presents how beliefs are collected. As shown in the table, we provide subjects with an incentive to truthfully report their belief: each option in Table 1 is associated with a payoff, computed via a quadratic scoring rule, which depends on a randomly selected choice, out of the three made, by the dictator with whom each subject is matched at end of the session. Subjects are informed that the relevant dictator for the payment of belief's accuracy cannot be the same dictator whose choice is selected for the payment of the dictator game. This way, payoff consequences of a dictator's choice do not extend to the guessing task but are limited to the game. Finally, to inhibit the updating of beliefs about restore probabilities by recipients, we avoid giving feedback during the three repetitions of the game.

Table 1: Elicitation of First–Order Beliefs

	Dictator will choose *Return*...			
	0 out of 3	1 out of 3	2 out of 3	3 out of 3
Your guess...	□	□	□	□
Your earnings if in the drawn choice...				
Dictator chose *Return*	€0	€2.80	€4.40	€5
Dictator chose *Keep*	€5	€4.40	€2.80	€0

In the instructions, we transparently inform group A members that their beliefs may be reported to dictators, to avoid a methodologically questionable omission of relevant information. A possible caveat of this design is that members of group A strategically manipulate their beliefs, knowing that with 50% probability their beliefs are revealed to dictators in Stage 3. However, we can control this issue *ex-post*, by comparing the beliefs distribution of group A with the beliefs distribution of group B, who know that their belief will not be disclosed to anyone.

[12] A similar solution to false consensus effect is used by Ellingsen et al. (2010) and Reuben et al. (2009), with the former finding no correlation between decision-maker choices and counterparts' expectations while the latter found evidence of guilt aversion.

4.1 Procedures

The experiment was conducted in the Cognitive and Experimental Economics Laboratory (CEEL) at the University of Trento. The experimental sessions were programmed and conducted using z-Tree software (Fischbacher, 2007). We conducted a total of 12 sessions and a total of 180 participants took part in only one session of the experiment. Subjects were recruited using dedicated software. All subjects received a show–up fee of €3.

As subjects entered the laboratory, they were randomly assigned to computer stations separated by partitions. In order to avoid any interaction between group A and group B, members of group B were asked to show up 10 minutes later than members of group A.[13] Moreover, the two groups were sitting at the opposite ends of the room, and those in group A were allowed to leave only after the payment of those in group B. Instructions were sequentially read aloud by the experimenter. Following Bigoni and Dragone (2012), as an aid to the instructions, we provided subjects with intuitive slides and we verified comprehension using a quiz.

5 Results

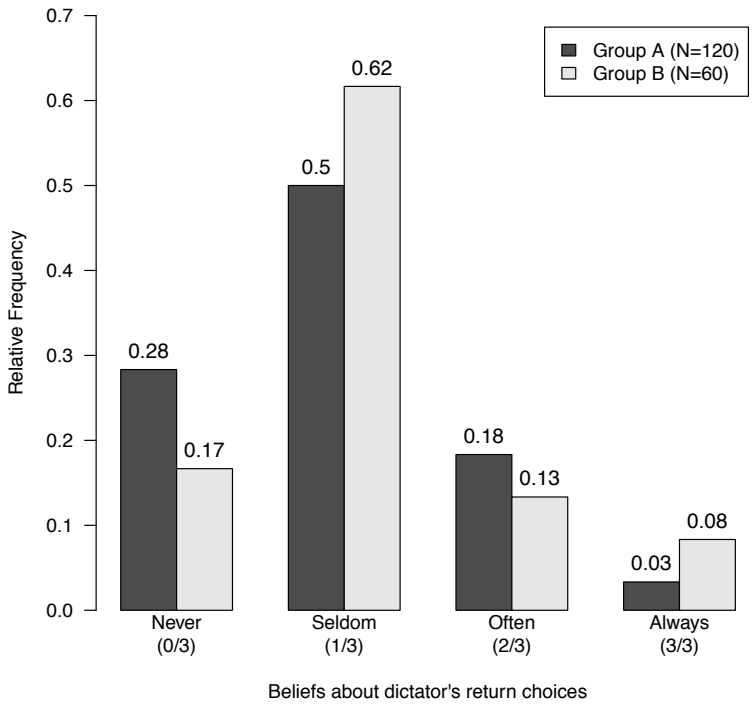
5.1 Descriptive Statistics

Figure 3 shows the separate distribution of the first-order beliefs about dictators' return choices for the A and B groups.[14] When self-reporting their beliefs, individuals in the A group are aware that this piece of information could be potentially disclosed to dictators. On the other hand, individuals in the B group know that their beliefs are not reported to dictators. Thus, the comparison between the self-reported beliefs of the two groups provides us with a control of the potential strategic manipulation of beliefs by prospective recipients. We adopt the labels "Never", "Seldom", "Often", and "Always" to identify beliefs that range from 0 returns out of 3 choices made, to 3 returns out of 3 choices made.

[13] In the experiment we decided to adopt colors to identify the two groups to ease the understanding of instructions. The "Green" label identified participants in group A and the "Red" label identified those in group B. In the instructions we also adopted the colors to graphically identify participants in the two groups.

[14] When beliefs are elicited, members of group A do not already know to which role they will be assigned in the modified dictator game.

Figure 3: First-Order Beliefs about Dictator's Return

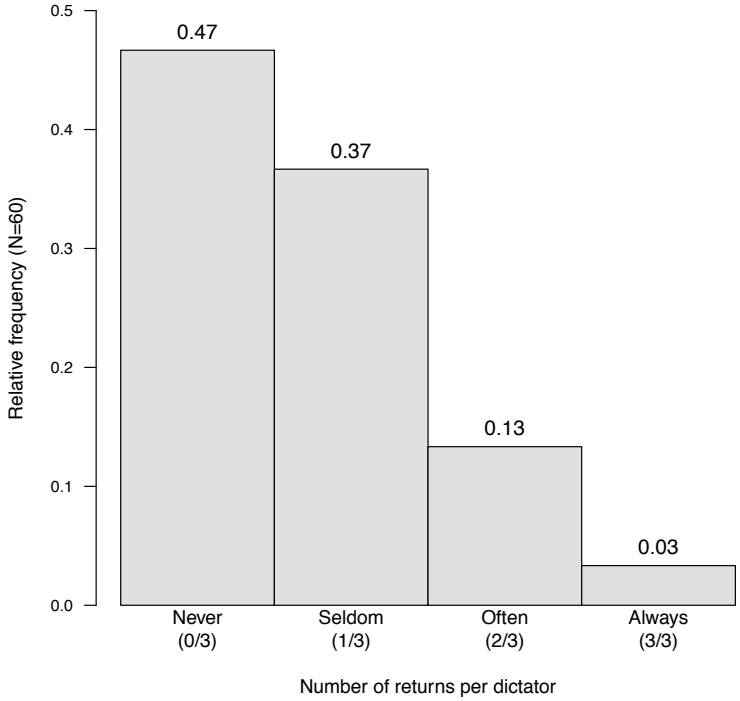


In both group A and B, a large majority of the participants (about 80%) expects a generic dictator to return the wallet less than 2 times out of the 3 total choices, with the mode of the distribution corresponding to the "Seldom" return frequency (1 out of 3). Although members of group A are slightly more pessimistic than group B members (0.32 vs. 0.37 respectively), when comparing the distributions of beliefs in the two groups, no statistically significant difference is observed (Fisher's Exact Test, p-value=0.117). Thus, we can safely conclude that those in the A group do not strategically manipulate their beliefs in a significant way.

Earnings in the belief task provide us with a direct measure of belief accuracy. Median earnings in the task are equal to €4.40 both for the red and the green group, just one step away from the maximum earnings of €5. Thus, self-reported beliefs seem to be overall quite accurate.

Figure 4 displays the relative frequency of individual-level return choices.

Figure 4: Return Choices per Dictator



As shown above, 47.0% of the dictators choose to never return. Among those choosing to return, the majority chooses to return the wallet only once and only 2 dictators out of 60 choose to always return the wallet. Overall, half of the dictators choose differently across the three rounds.

Figure 4 clearly shows that about half of collect choices are compatible with selfishness. Concerning the conspicuous share of non-selfish dictators, very few choices are compatible with standard outcome-based social preferences. As pointed out in Section 3, non-selfish individuals concerned only with the payoff consequences of their actions are likely to always return the wallet, irrespective of the restoring probability and of beliefs of the counterpart. The behavior of non-selfish dictators seems to be largely conditional upon the contextual elements in which the choice is made for different probabilities of success, with more dictators returning for the highest probability (30.0%) than for the other two probabilities of success (21.7%). Table 2 reports the percentage returns observed for alternative levels of restore probability and of counterpart's beliefs, both when considered alone and when jointly taken. As an example, the first cell in the table shows that out of the 19 dictators who jointly faced a probability of success of 4/6 and a recipient's belief of zero returns in a round, only 1 of them decided to return the wallet (i.e. 5.3%).

As shown by the last column in Table 2 (*Overall*), in correspondence to the highest

Table 2: Return Choices conditional upon Recipient's Beliefs and Restoring Probabilities

	Recipient's First-Order Belief				
PrRestore	Never (N=19)	Seldom (N=27)	Often (N=13)	Always (N=1)	*Overall*
4/6	5.3%	25.9%	38.5%	0.0%	21.7%
5/6	15.8%	25.9%	23.1%	0.0%	21.7%
6/6	31.6%	37.0%	15.4%	0.0%	30.0%
Overall	17.5%	29.6%	25.6%	0.0%	24.4%

probability of restore we can observe the highest percentage in returns, which is in line with Hypothesis 1. However, the percentage of returns in the intermediate probability of restore and in the lowest do not differ. The bottom row in Table 2 provides evidence about the impact of a counterpart's beliefs and shows that the lower percentage of returns is observed in correspondence to the lowest level of beliefs, in line with Hypothesis 2, but the maximum share of returns is observed in correspondence to "Seldom" rather than to "Often".[15]

When taking into account the joint effect of beliefs and restore probabilities, the frequency of returns monotonically increases in the beliefs of the counterpart, in correspondence to $PrRestore = 4/6$. A Fisher's exact test (FET) shows that there is a statistically significant difference between return choices in correspondence to the "Never" and "Often" belief levels (p-value= 0.029). Moreover, the return rates observed in correspondence to these two belief levels are the overall lowest and highest respectively. In contrast, for $PrRestore = 5/6$ and $PrRestore = 6/6$ the impact of beliefs of the counterpart seems to be more erratic, with no significant differences in return choices in correspondence to the "Never" and "Often" belief levels (FET, p-values$\geq$ 0.420).[16]

The descriptive analysis reported above suggests that both restore probabilities and beliefs of the counterpart have a positive impact on return decisions even though the impact of these variable is not fully in line with Hypothesis 1 and 2. Moreover, beliefs and probabilities seem to interact in shaping dictator choices in a way that is in contrast to our Hypothesis 3. Below we report a regression analysis that provides us with a further

[15]The belief level "Always" is neglected in the discussion because only one observation is available.

[16]For any $PrRestore$, the comparisons of "Never" and "Seldom" and of "Seldom" and "Often" do not reveal any significant difference (all p-values$\geq$ 0.115)

test of our three main hypotheses and which casts further light on the behavior in the experiment.

5.2 Regression Analysis

Table 3 reports the outcomes of a series of Generalized Linear Mixed Logit (GLM Logit) models controlling for repeated choices at the individual level through random effects. The dependent variable in all regressions is the dictator's decision to return the wallet (*Return*), assuming value 1 when the dictator returns and 0 otherwise.

Four distinct estimates are presented in Table 3. In Model 1, we only consider first-order beliefs of the recipient (*RecBel*) as the main explanatory variable, with *RecBel* assuming values $\{\frac{0}{3}, \frac{1}{3}, \frac{2}{3}, \frac{3}{3}\}$ for beliefs equal to "Never", "Seldom", "Often", and "Always", respectively. In Model 2, we also control for the alternative probabilities with which the entitled recipient restores her wallet once the dictator chose to return it ($PrRestore$), i.e. $\frac{4}{6}, \frac{5}{6}$, or $\frac{6}{6}$. In Model 3, we also include the interaction between recipients' beliefs and probability of success ($RecBel \times PrRestore$). The main explanatory variables of Model 3 refer explicitly to our main research hypotheses on conditional guilt (Hypothesis 3), testing for the impact of beliefs and probabilities both in isolation and in interaction. Additionally, Model 4 introduces, as an explanatory variable, the dictator's first-order belief about the behavior of other dictators (*DictBel*) and how these belief stand relative to the belief of the recipient, with a specific focus on the case in which beliefs of the recipient are more optimistic than those of the dictator ($RecBel > DictBel$).[17]

Finally, for each estimate we considered a set of control variables regarding gender ($Male$), employment status ($Worker$), citizenship ($Italian$), and previous participation in experiments ($Experienced$).

[17]The beliefs of the dictator are collected before knowing the actual role in the game. Thus, *DictBel* and *RecBel* refer to the same set of beliefs.

Table 3: Determinants of return choices.

Dep. Var.: Return (0,1)	Model 1	Model 2	Model 3	Model 4
(Intercept)	−0.750 (0.712)	−2.040 (1.346)	−5.555 (2.111)**	−5.899 (2.148)**
RecBel	0.695 (0.734)	0.718 (0.739)	11.909 (4.877)*	13.028 (5.344)*
PrRestore		1.531 (1.341)	5.880 (2.367)*	5.544 (2.417)*
RecBel × *PrRestore*			−13.287 (5.679)*	−12.824 (6.136)*
DictBel				2.559 (0.961)**
RecBel > *DictBel*				−1.612 (0.747)*
Male	−0.289 (0.412)	−0.290 (0.417)	−0.484 (0.411)	−0.471 (0.405)
Worker	−0.019 (0.223)	−0.019 (0.226)	−0.058 (0.217)	−0.110 (0.220)
Italian	−0.865 (0.581)	−0.877 (0.591)	−0.964 (0.567)°	−0.820 (0.538)
Experienced	0.248 (0.501)	0.247 (0.508)	0.330 (0.490)	−0.270 (0.512)
AIC	209.519	210.241	206.721	184.846
BIC	231.870	235.784	235.457	219.968
Random effects (sbj. ID)	*Yes*	*Yes*	*Yes*	*Yes*
Num. obs.	180	180	180	180
Num. groups: ID	60	60	60	60

Note: Standard errors in parentheses. All estimates are from GLM Logit models including random effects at subject-level. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\circ}p < 0.1$.

The information criteria reported in the lower panel of Table 3 (AIC and BIC) unanimously suggest that Model 4 best exploits the information available and, thus, we focus on this specification when commenting on our results. However, as shown by the comparison of Models 3 and 4, the main results are robust across alternative specifications.[18]

Model 4 shows that a higher probability of restoring the wallet to the entitled recipient induces more returns on the side of dictators, providing support to the hypothesis of justice concern (Hypothesis 1). In other words, when the probability of a misplaced return is lower, dictators are more likely to choose to return. The positive and statistically significant coefficient of *RecBel* suggests that more optimistic beliefs of the recipient are more likely to trigger a return than less optimistic ones. This provides support to the hypothesis of guilt aversion (Hypothesis 2). The estimated coefficient of the interaction between recipient's beliefs and probability of success ($RecBel \times PrRestore$) is negative. Thus, in contrast to our hypothesis of conditional guilt (Hypothesis 3), facing recipients with more optimistic beliefs has a stronger positive impact on the decision to return when restore probabilities are lower.

[18]Results are robust to specifications where controls and/or random effects are omitted.

Finally, controls on dictator's first-order beliefs suggest that expectations about behavior of others in the same role are positively correlated to dictators' choices ($DictBel$). Furthermore, when the beliefs of the recipient are too optimistic relative to those of the dictator ($RecBel > DictBel$), returns are less likely.

6 Discussion and Conclusions

Results from our experiment demonstrate that justice concerns, originating in the balance between effort and rewards, play an important role in shaping dictators' choices in our modified dictator game. When there is a higher probability that the endowment is returned to the individual who did not work to generate it, fewer dictators decide to return. This result hence adds to previous works highlighting the relevance of the proportionality between inputs and outputs in a production process (e.g. Adams, 1963; Konow, 2000).

Our experiment also shows that dictators react to the expectations of their counterpart, by being more likely to return when facing optimistic recipients. As a result of our chosen experimental design, it is possible to causally interpret the effect that recipients' beliefs have on dictators choices. After ruling out confounds as the false *consensus effect* and verifying the absence of strategic manipulation of expectations, we can safely interpret this finding as an attempt of dictators to avoid the negative emotion of guilt. While evidence on guilt aversion is still mixed, our findings are partially consistent with those experiments providing empirical support to this theory.

Our results clearly show that both justice concerns and counterpart's expectations affect dictators. We further contribute to the literature by showing how these two relevant sources of behavior interact. The intuition driving our research hypothesis is that counterpart's expectations are more likely to affect behavior when they do not conflict with justice considerations. To our surprise, however, the expectations of the entitled recipients are more likely to be fulfilled when the likelihood of a misplaced return is higher. In other words, a higher protection of entitlement rights crowds out the role of others' expectations. We believe that this result may deserve further attention as it provides interesting insights into the working of institutions, which are proxied by different restoring probabilities in our experiment. Data collected suggest that when institutions are weaker, i.e. they offer less protection to legitimate property rights, individuals may rely more on a subjective measures of justice, as captured by the expectations of the counterpart which

trigger a sense of guilt when disappointed.

Finally, we observe that, together with beliefs of the counterpart, the beliefs of the decision-maker also play a crucial role in explaining behavior. We hence attempt to rationalize the evidence from Model 4 in Table 3 in the light of the theoretical framework proposed by Bicchieri (2006), according to which individuals are more inclined to comply to a norm when, (1) they expect others to follow it, and (2) they are expected by others to follow it, conditional upon others' expectations to be legitimate. In line with the first part of Bicchieri's argument, we find a strong positive correlation between dictators' beliefs about a generic dictator choosing to return the wallet and their own decision to return. At the same time, we see that dictators facing over-optimistic beliefs (with respect to their own) tend to react in the opposite way, suggesting that for others' expectations to be effective these must appear appropriate.

References

Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *The American Economic Review*, 101(2):470–492.

Adams, S. J. (1963). Toward an Understanding of Inequity. *Journal of Abnormal and Social Psychology*, 67(5):422–436.

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

Battigalli, P., Charness, G., and Dufwenberg, M. (2013). Deception: The role of guilt. *Journal of Economic Behavior & Organization*. Forthcoming.

Battigalli, P. and Dufwenberg, M. (2007). Guilt in games. *The American Economic Review: Papers and Proceedings*, 97(2):170–176.

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

Becker, A. (2013). Accountability and the fairness bias: the effects of effort vs. luck. *Social Choice and Welfare*, 41(3):685–699.

Bellemare, C., Sebald, A., and Strobel, M. (2011). Measuring the willingness to pay to avoid guilt: estimation using equilibrium and stated belief models. *Journal of Applied Econometrics*, 26(3):437–453.

Bicchieri, C. (2006). *The grammar of society: the nature and dynamics of social norms*. Cambridge University Press, New York.

Bigoni, M. and Dragone, D. (2012). Effective and efficient experimental instructions. *Economics Letters*, 117(2):460–463.

Bolton, G. E. and Ockenfels, A. (2000). Erc: A theory of equity, reciprocity and competition. *The American Economic Review*, 90(1):166–193.

Cappelen, A. W., Hole, A. D., Sorensen, E. O., and Tungoddedn, B. (2007). The pluralism of fairness ideals: An experimental approach. *The American Economic Review*, 97(3):818–827.

Cervellati, M., Esteban, J., and Kranich, L. (2000). Work values, endogenous sentiments redistribution. *Journal of Public Economics*, 94(9):612–627.

Charness, G. and Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, 74(6):1579–1601.

Charness, G. and Dufwenberg, M. (2010). Bare promises: An experiment. *Economics Letters*, 107(2):281–283.

Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 117(3):817–869.

Cherry, T. L., Frykblom, P., and Shogren, J. F. (2002). Hardnose the dictator. *The American Economic Review*, 92(4):1218–1221.

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6):1015–1026.

Cox, J. C., Friedman, D., and Sadiraj, V. (2008). Revealed altruism. *Econometrica*, 76(1):31–69.

Dufwenberg, M., Gächter, S., and Hennig-Schmidt, H. (2011). The framing of games and the psychology of play. *Games and Economic Behavior*, 73(2):459–478.

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

Ellingsen, T. and Johannesson, M. (2004). Promises, threats and fairness. *The Economic Journal*, 114(495):397–420.

Ellingsen, T., Johannesson, M., Tjøtta, S., and Torsvik, G. (2010). Testing guilt aversion. *Games and Economic Behavior*, 68(1):95–107.

Elster, J. (1998). Emotions and economic theory. *Journal of Economic Literature*, 36(1):47–74.

Engelmann, D. and Strobel, M. (2000). The false consensus effect disappears if representative information and monetary incentives are given. *Experimental Economics*, 3(3):241–260.

Engelmann, D. and Strobel, M. (2011). Deconstruction and reconstruction of an anomaly. *Games and Economic Behavior*, 76(2):678–689.

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

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

Geanakoplos, J., Pearce, D., and Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior*, 1(1):60–79.

Gneezy, U. (2005). Deception: The role of consequences. *The American Economic Review*, 95(1):384–394.

Hoffman, E., McCabe, K., Shachat, K., & Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior*, 7(3): 346–380.

Hoffman, E. and Spitzer, M. L. (1985). Entitlements, rights, and fairness: An experimental examination of subjects' concepts of distributive justice. *The Journal of Legal Studies*, pages 259–297.

Kartik, N. (2009). Strategic communication with lying costs. *Review of Economic Studies*, 76(4):1359–1395.

Konow, J. (1996). A positive theory of economic fairness. *Journal of Economic Behavior & Organization*, 31(1):13–35.

Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. *The American Economic Review*, 90(4):1072–1091.

Konow, J. (2003). Which is the fairest one of all? a positive analysis of justice theories. *Journal of Economic Literature*, XLI:1188–1239.

Leventhal, G. S. and Michaels, J. W. (1969). Extending the equity model: Perceptions of inputs and allocation of reward as a function of duration and quantity of performance. *Journal of Personality and Social Psychology*, 12(4):303–309.

Lindbeck, A. and Nyberg, S. (2000). Raising children to work hard: altruism, work norms, and social insurance. *The Quarterly Journal of Economics*, 121(4):1473–1503.

Locke, J. (2011). *Second Treatise of Government*. Simon & Brown.

Mikula, G. (1974). Nationality, performance, and sex as determinants of reward allocation. *Journal of Personality and Social Psychology*, (4):425–440.

Ockenfels, A. and Werner, P. (2014). Scale manipulation in dictator games. *Journal of Economic Behavior & Organization*, 97:138–142.

Reuben, E., Sapienza, P., and Zingales, L. (2009). Is mistrust self-fulfilling? *Economics Letters*, 104(2):89–91.

Tangney, J. (1995). Recent Advances in the Empirical-Study of Shame and Guilt. *American Behavioral Science*, 38(8):1132–1145.

Vanberg, C. (2008). Why do people keep their promises? an experimental test of two explanations. *Econometrica*, 76(6):1467–1480.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological review*, 92(4):548.

A Instructions

[ONLY THE FIRST 10 SUBJECTS ARE IN THE LAB]

Welcome!

For showing up you have earned €3. During the session you can earn more money. The entire amount will be paid to you in private at the end of this session. Please, follow the instructions carefully and do not speak to the other participants. If you have questions, raise your hand and one of the experimenters will answer to you in private.

The session consists of 4 stages. You are identified as the *Green Group* and you are going to complete the first stage. After the first stage, another group of participants will enter this room,and will be identified as the *Red Group*.

Stage 1 – Work. In this stage you have to count the number of zeros within seven tables that will consecutively appear on your screen. For each table you will earn 1 token once you have inputed the correct number of zeros.

In a following stage, something unexpected could occur which implies the loss of the tokens you have earned. At the end of the session, each token you own will be converted into €1.

The Red Group can now come in.

[RED GROUP ENTERS THE LAB]

Welcome!

For showing up you have earned €3. During the session you can earn more money. Please, follow the instructions carefully and do not speak to the other participants.

The 10 participants who were already in this room are identified as the *Green Group*. The 5 participants that have just entered are instead identified as the *Red Group*.

The members of the Green Group have already been here for about 15 minutes, and have just completed Stage 1 of the session, which consisted of counting the number of zeros in a sequence of tables. For doing the work, every member of the Green Group has earned 7 tokens. At the end of the session, each token you own will be converted into €1.

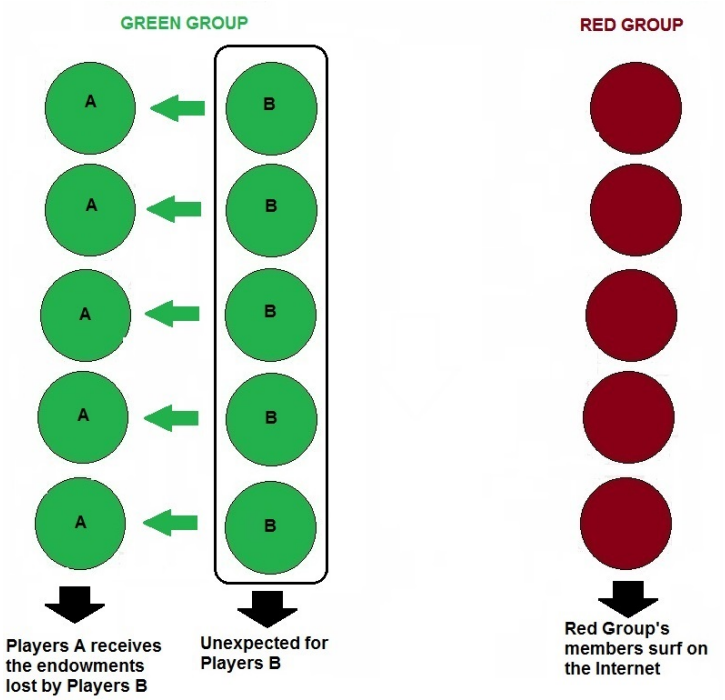
The members of the Red Group will participate only in the next stage of the session, and then they will be allowed to freely surf on the Internet, but the members of the Green

Group will have to participate actively in the whole session. Before being allowed to leave, they will have to wait until all the members of the Red Group are paid.

To understand the Stage 2 (*estimation*) you need to know about Stage 3 (*unexpected*) and Stage 4 (*decision*). Before proceeding to Stage 2, we will give you the instructions for stage 3 and 4. *Everyone,* please carefully follow the instructions.

Stage 3 – Unexpected. In this stage none of the participants are asked to take any decisions. The computer will execute all the procedures automatically.

If you are a member of the Green Group, you could suffer an *unexpected*: one person every two, randomly selected by the computer, will lose the 7 tokens earned from the work in Stage 1 and will become *Participant B* (see the picture). The lost earnings of Participant B will be given to another person, *Participant A*, randomly selected from among the 5 members of the Green Group who did not suffer the unexpected (see the picture).



If you are a Green Group member, on your screen you will see the role that was assigned to you. If you are a Red Group member, you will be allowed to surf the Internet.

Stage 4 – Decision. Every Participant A is paired with the Participant B from whom he/she has received the gain. Each Participant A will have to decide whether to return the gain to Participant B:

- If he/she decides to *not return*, A earns a total of 12 tokens and B remains with 0 tokens.

- If he/she decides to *return*, A earns a total of 9 tokens and B restores his/her initial earning of 7 coins with a probability between 67% and 100%. In the cases where B does not restore his/her earnings, the 7 tokens are transferred to a member of the Red Group, selected at random. The outcome of the restitution is determined with a die roll (performed by the computer).

Participants' total earnings are summarized in the following diagram:
FIGURE HERE

The situation just described will be repeated three times. In each repetition, each Participant B will be paired again with a new Participant A, different from the one previously met (i.e. you will never be paired with the same person for more than one repetition). Are there any questions about this?

There are still three important things to say about Stage 4:

1. In each of the three repetitions, the recovering probability is different and randomly changes between 67% and 100% in the following ways:

   Important: Only Participant A knows the exact probability with which the 7 tokens will be recovered by B. Participant B will *never* know, because the order with which the different recovering probabilities appear is random and remains unknown to B.

2. All the decisions will be taken sequentially, without receiving feedback on their outcome. At the end of Stage 4, every participants will see the number of tokens earned. Only Participant A will be informed about the outcome of his/her choice in the repetition randomly selected for the payment (see below).

3. One of the three repetitions, selected at random by the computer, will be paid at the end of the session.

Before proceeding, we will ask you to answer some control questions about the instructions.

We can now return to Stage 2.

Stage 2 – Estimation. You are now asked to report what decisions you expect from Participants A in Stage 4 (decision). In particular, you have to estimate how many times you expect a generic Participant A will choose to return the gain to the Participant B across the three repetitions (from 0 out of 3, to 3 out of 3). Is that clear?

The more your estimate is accurate, the more money you can earn. To evalauate the accuracy of your estimate, the computer will match you to a Participant A selected at random. The Participant A randomly selected to determine your earning for the Stage 2 (estimate) will be different from the Participant A who will be selected at random to determine your earning in the stage 4 (decision). At the end of the session, the computer will draw at random one of the three decisions made by Participant A you are matched to. The computer will compare your estimate with the choice of A in the drawn decision (the choice could be either "Return" or "Not return").

In the table you can see how much you will earn given your estimate and the choice of Participant A.

	Dictator will choose *Return*...			
	0 out of 3	1 out of 3	2 out of 3	3 out of 3
Your guess...	□	□	□	□
Your earnings if in the drawn choice...				
Dictator chose *Return*	0 €	2.80 €	4.40 €	5 €
Dictator chose *Keep*	5 €	4.40 €	2.80 €	0 €

Note: During Stage 4, before each decision, Partecipants A will be also informed about the estimate made in Stage 2 by Participant B with whom they are paired.

Payment: The total amout paid at the end of the session (in addition to the €3 for showing up) is the sum of the gain for the accuracy of your estimate in Stage 2, and your earnings in the repetition randomly selected from Stage 4.

From here on, Red Group members can surf the Internet.