Laboratory tests of theories of strategic interaction

Yang, Y.

Publication date
2014
Document Version
Final published version

3

Is Ignorance Bliss?

3.1 Introduction

Is it beneficial to know the kind of person you are dealing with in a social dilemma situation? Does it hurt you, if your type is revealed to someone else? These questions are at the heart of many debates on public policies. For example, many U.S. states allow juvenile court convictions to be expunged.[1] As a consequence, a former offenders' future employer will not find her or her juvenile criminal records in a background check in a recruitment procedure. Also, in many countries individual credit records are cleared after a certain amount of time.[2] Allegedly, this is to give the individuals concerned a 'fresh start'. It therefore seems to benefit the 'bad' types. It may however, harm the 'good' types (being unable to distinguish themselves from the 'bad') and possibly also harm those who interact with such individuals without knowing their past record (like a future employer). Whether the aggregate effect is positive or negative is an open question. In this chapter, we address this question and investigate the consequences of type revelation in a binary social dilemma.

We first present a model and provide a game-theoretical analysis. Our model involves a two-round prisoner's dilemma with random rematching between the two rounds. Players' choices in the first round can be observed by their round-2 opponent. We introduce two types of players, 'Givers' and 'Takers', who have different preference levels for inequity aversion (Fehr and Schmidt 1999). The former have intrinsic preferences for cooperation while the latter do not. Finally, our model distinguished between two type information mechanisms: in one, player types are revealed before actions are chosen. In the other, types remain private information.

Cooperation in a repeated prisoner's dilemma game with random rematching has been widely studied (e.g., Ellison 1994). Typically, more cooperation is observed in such 'strangers' environments than would be predicted by standard economic models based on selfish preferences. Cooperation is further fostered (and may be part of an equilibrium) in repeated interactions between fixed partners where private monitoring of the opponent's history is possible (e.g. Fudenberg and Maskin 1986). In our model, we introduce limited monitoring of past choices after rematching. The assumption that players can observe their opponent's past action towards a third

[1] As an example, the relevant legislation in Utah can be found at `http://www.judiciary.state.nj.us/prose/10557_expunge_kit.pdf`.

[2] See `http://www.equifax.com/answers/request-free-credit-report/en_efx`.

player is based on the real-life observation of the existence of reputation mechanisms (e.g. online trading reputation records), which facilitates access to a trading partner's history. We will show that the introduction of reputational information in a strangers design boosts cooperation levels. The main interest in this chapter, however, lies in how such information about reputation interacts with information about an individual's type. Note that an individual's type will in some sense be more informative about behavior to expect than previous choices, since such choices may be made strategically, with future interactions in mind.

On the other hand, such strategic choices may be desirable in the sense that they yield higher cooperation levels in early rounds. The perfect Bayesian equilibrium (PBE) of our model indeed predicts that in certain environments, not knowing the type of person you are dealing with may result in higher cooperation rates and hence be beneficial to all. This is because it eliminates the discrimination of Takers and gives them incentives to behave cooperatively for strategic reasons. This result is reminiscent of the finding by Hu and Qin (2013), who find that efficiency in a financial market is decreased after revelation of information.

However, the observations from our laboratory experiment do not support the theoretical prediction. Givers and Takers both exhibit higher cooperation rates when type information is revealed. Moreover, both types are found to reciprocate positively opponents who behaved cooperatively towards others in the past, even when such a behavior is suboptimal to Takers. This so-called 'indirect reciprocity' is found to be stronger when type information is revealed, than in the un-revealed treatment, which contributes to the increase in the cooperation rates when revealing the subjects' types to their opponents.

Another interesting observation from or data is that, even though in general, subjects engage in indirect reciprocity, they do not fully strategically respond to others' indirect reciprocity. Such strategic response can be expected from players who could do as little as one step of strategic thinking. This provides a further explanation of why the prediction provided by the game-theoretical analysis, which is based on the assumption of agents' full rational and strategic thinking, is not supported in our experiment.

In an attempt to better understand our results, we will fit our data to an image scoring model, as introduced by Nowak and Sigmund (1998) in an evolutionary game framework, and further studied by Wedekind and Milinski (2000) and Seinen and Schram (2006) in laboratory experiments. In contrast to the game-theoretical

model, the image scoring model requires no strategic thinking. Instead, it assumes that players follow a given decision rule, based on the observed reputation of the people they interact with. The estimated decision rule for Takers corresponds to their predicted strategy in the one-shot prisoner's dilemma game. For Givers, their decision rules estimated using this model is more cooperative than predicted by PBE in the game-theoretical model. In aggregate, about 60% and 70%, respectively, of the Givers and Takers' decisions are correctly predicted by the image scoring model.

The remainder of this chapter is structured as follows. We introduce the game-theoretical model and provide the equilibrium analysis in Section 3.2. Section 3.3 presents our experimental design and procedures. The discussion of the results is presented in Section 3.4. Section 3.5 summarizes our findings.

3.2 Model

3.2.1 Primitives

	Give (G)	Take (T)
Give (G)	c, c	f, d
Take (T)	d, f	s, s

Game 1: Pecuniary Payoffs of the Stage Game

We consider a two-round game, where a simultaneous two-player game is played in each round. In the stage game, each player can choose between *Give* (G) or *Take* (T). The monetary payoff structure of the stage game is as shown in Game 1, where $d > c > s > f$ and $2c > f + d$. In terms of the monetary payoffs, the stage game is a prisoner's dilemma, although it need not be so in terms of players' utility, as we will discuss in details later. Hence we will refer to the stage game and the two-round game, respectively, by 'PD' and '2PD' in the remainder of this chapter.

Player Heterogeneity

$$u(x, y) = x - \alpha \max(y - x, 0) - \beta \max(x - y, 0). \tag{3.1}$$

The model distinguished between two types of players, 'Giver' and 'Taker' (the motivation underlying this choice of labels will become clear shortly). The two types are distinguished by their distinct attitudes towards fairness, i.e. the envy (α) and guilt (β) parameters in the inequality aversion (IA) model introduced by Fehr and Schmidt (1999) (eq (3.1)). While Takers care only about their own monetary payoffs ($\alpha^T = \beta^T = 0$), Givers are characterized by inequity aversion, with envy and guilt parameters α^G and β^G. We assume that these IA parameters satisfy:

$$c > (1 - \beta^G)d + \beta^G f \tag{3.2}$$

condition implies that for a Giver, the best response to *Give* in Game 1, is *Give*. Hence, when two Givers meet in Game 1, (G, G) is an equilibrium (as is $(T,\ T)$). Moreover, $(G,\ G)$ will be the preferred Nash equilibrium for both players. In contrast, for a Taker, T is the dominant strategy. Hence, independent of the player types, $(T,\ T)$ is always a Nash equilibrium while $(G,\ G)$ is always the socially efficient outcome. The latter is only an equilibrium when two Givers meet, however.

Random Matching At the beginning of each round of the 2PD game, players are randomly paired. Therefore, there is no repeated interaction. Even though there is a small chance that two players remain in a pair in both rounds, they cannot recognize each other and therefore cannot know that their interaction has been repeated.

Information Structure In this model, we consider two different information mechanisms.

The two mechanisms differ in whether players' types are revealed. In the first mechanism, each player's type is revealed to her current opponent in each round before decisions are made. In the second mechanism, each player's type information remains private and is not revealed to any other player. The two information conditions share two common features. First, at the beginning of Round 2 before each player makes a decision, she is informed about the action chosen by her current opponent Round 1. Second, every player always knows her own type and the fraction of Givers in the population, denoted by ρ (so the fraction of Takers, $1 - \rho$, is also commonly known).

3.2.2 Equilibrium

In this section, we present the PBE for the two information mechanisms in section 3.2.1. Then, based on the PBE under each information mechanism, we give a theoretical prediction of the comparison between the cooperation rates under the two mechanisms.

When the types are revealed, a player's strategy in a symmetric PBE can be written in the following general form: $S^{t_i} = (s_1^{t_i}(t_{-i_1}), s_2^{t_i}(t_{-i_2}, a_1^i, a_1^{-i_2}))$, where $s_m^{t_i}$ is the choice for round m=1,2, $t_i \in \{Giver, Taker\}$ denotes player i's type, $a_1^i \in \{G, T\}$ is the (revealed) choice made in Round 1 by player i and $-i_1, -i_2$ denote the index of player i's paired opponent in Round 1 and 2, respectively.

Lemma 1. *If the type is revealed, for any fraction of the Givers (i.e., $\forall \rho \in [0,1]$), the symmetric PBE with the highest cooperation rate is*

$$s_1^{Giver}(t_{-i_1}) = \begin{cases} G & \text{if } t_{-i_1} = Giver, \\ T & \text{if } t_{-i_1} = Taker; \end{cases} \tag{3.3}$$

$$s_2^{Giver}(t_{-i_2}, a_1^i, a_1^{-i_2}) = \begin{cases} G & \text{if } t_{-i_2} = Giver, \\ T & \text{if } t_{-i_2} = Taker. \end{cases} \tag{3.4}$$

$$s_1^{Taker}(t_{-i_1}) = T, \quad \forall t_{-i_1} \in \{Giver, Taker\}; \tag{3.5}$$

$$s_2^{Taker}(t_{-i_2}, a_1^i, a_1^{-i_2}) = T, \quad \forall (t_{-i_2}, a_1^i, a_1^{-i_2}) \in \{Giver, Taker\} \times \{G, T\}^2. \tag{3.6}$$

The idea underlying this perfect Bayesian equilibrium is very intuitive and also serves as the proof for the Lemma. First, because T is the dominant strategy for Takers, in round 2 (the final round) Takers will always choose T, regardless of the opponent's type, round-1 choice, or the own choice in the previous round. Anticipating this, any Giver that is matched with a Taker will choose the best response to T, which is also T, no matter what the opponent or she herself has chosen in round 1.

Then by backward induction, in Round 1, knowing that whatever they choose will have no influence on the next-round outcome, a Taker will stick to their stage-game dominant strategy T, and Givers will still always choose T against any Taker they are matched with. On the other hand, because Givers can always recognize

each other when matched, they can always realize the stage equilibrium, which is also the socially efficient outcome, (G, G), in both rounds, regardless of information on previous choices. These strategies form a PBE of the 2PD, since no one has an incentive to deviate. Also, these strategies achieve the highest cooperation rate, since all possible cooperation (which can only be achieved between Givers) is realized. The above serves as the proof of Lemma 1.

Under the mechanism where the type information is not revealed, in a symmetric PBE, a player's strategy can be represented by $S^{t_i} = (s_1^{t_i}; s_2^{t_i}(a_1^i, a_1^{-i_2}))$, where $t_i \in \{Giver, Taker\}$ denotes player i's type; $s_1^{t_i} \in \{G, T\}$ is player i's choice in Round 1 given her type as t_i, and $s_2^{t_i}(a_1^i, a_1^{-i_2})$ is the choice of a type-t player's choice in Round 2, which is a function dependent on the (revealed) choices in the previous round of the player herself and her Round-2 opponent.

Lemma 2. *In the 2PD under the type-information-hiding information mechanism, the following strategies construct a PBE*

$$s_1^{Giver} = G; s_2^{Giver}(a_1^i, a_1^{-i_2}) = \begin{cases} G & \text{if } a_1^i = a_1^{-i_2} = G, \\ T & \text{otherwise.} \end{cases} \tag{3.7}$$

$$s_1^{Taker} = G; s_2^{Taker}(a_1^i, a_1^{-i_2}) = T, \forall a_1^i, a_1^{-i_2} \in \{G, T\}, \tag{3.8}$$

if

$$\rho \geq max\{\frac{d-c}{d-s}, \frac{s-f+\alpha^G(d-f)}{c+s-d-f+(\alpha^G+\beta^G)(d-f)}\}. \tag{3.9}$$

The intuition underlying Lemma 2 is that, when the fraction of the Givers in the population is high enough, in the second round it can be optimal for a Giver to choose G even if she is aware of the risk of being matched with a Taker (who chooses T in round 2 for sure) as long as the probability of such a match is low enough. Anticipating the chance of being matched with a Giver in round 2 who will discriminate based on round- 1 choices, Takers have sufficient incentive to build a good reputation by mimicking Givers and choosing the cooperative action G in round 1. A detailed proof of lemma 2 is provided in Appendix 3.6.1.

From Lemmas 1 and 2, we can directly obtain the following proposition.

Proposition 1. *In the 2PD game, if the fraction of Givers in the population satisfies*

$$\rho \geq max\{\frac{d-c}{d-s}, \frac{s-f+\alpha^G(d-f)}{c+s-d-f+(\alpha^G+\beta^G)(d-f)}\}, \tag{3.10}$$

then higher cooperation rates will be observed without type revelation than the best results that can be reached under the mechanism where players' types are revealed to their opponents.

Proposition 1 suggests that if the fraction of Givers in the population is sufficiently high, withholding information on players' types is more efficient than providing it, because leaving types hidden gives Takers an incentive to mimic the cooperative behaviour of the Givers in the early round. This leads to different choices by both Takers and Givers. We call the result that type information reduces efficiency the "Ignorance is Bliss" hypothesis.

A comparison of the predicted cooperation rates under the two information mechanisms shows that in Round 1 the increase in cooperation from withholding type information comes from two parts. The first is that all Takers switch from T to G and the second comes from the Givers who are paired with Takers (who choose T if they know the opponent is a Giver and G if they don not). Therefore, the increase in cooperation when types remain unknown is a fraction $1-\rho+\rho(1-\rho) = 1-\rho^2$ of the population. In Round 2, in comparison to the situation where the type is observable, more Givers choose G when types are kept hidden with the difference coming from those Givers who are paired with a Taker, which is proportion $\rho(1-\rho)$. In round 2, Takers choose T under both information systems. Hence when condition (3.10) holds, the frequency of players choosing G increases in both rounds.

3.3 Experiment

To test the theory we have designed a laboratory experiment consisting with two parts. The experiment starts with a first part used to identifying subjects' types by measuring their inequity aversion (i.e. their α and β levels); and subsequently tests subjects' behavior in the 2PD game (and its variations) in a second part. The instructions used in the experiment are included in Appendix 3.6.2.[3] We will start

[3]Before Part 1, we also asked the subjects to play an ultimatum game in a strategy method (everyone needed to submit a proposal on how to split 100 points within her pair as if she was to become the proposer, and also a threshold indicating the lowest amount for her to accept the

by explaining the menus used in Part 1.

Part 1: Preference Measurement

In this part, we implement two menu tests used in Yang, Onderstal, and Schram (2013) and Chapter 2. We present two menus (Menu 1 & 2 as shown in Tables 3.1 and 3.2) to the subjects. Each menu consists of a list of decisions for each subject to make as a **proposer**. For each decision, a proposer needs to choose between two payoff bundles given in Option A and B, which specify the payoff for herself and her **receiver** (another subject paired with her her for this part, who receives the amount that the proposer chooses for 'other' in the Menu). In Menu 1 (as shown in Table 3.1), for each decision, the amount (of experimental points) for the proposer is always lower than the amount for the receiver in both Options A and B. Therefore, for each decision, a choice of either A or B reflects the proposer's α level. The last column of Table 3.1 gives the lower bound on a proposer's α level such that A would be chosen for the respective decision.[4] This column is, of course, not displayed to the subjects in the experiment. Analogously, a choice in B for each decision implies an inequality in α with the opposite sign as shown in the table. Therefore, given a subject's α level, which is assumed to be non-negative, the subject's rational behavior for the decisions in Menu 1 should start with choosing A in decision 1 and sticking to A until she switches to B from one decision and there afterwards. For example, if a proposer has $\alpha = 0.10$, she would choose A for decision 1, 2 and switch to B from decision 3 onward. As having been discussed in Chapter 2, any subject whose utility follows eq (3.1) will switch from A to B at most once in Menu 1, and the switching point from A to B allows us to determine an interval for the subject's envy parameter.

In Menu 2 (displayed in Table 3.2) the payoff for the receiver is always less than the proposer's. Thus, according to eq (3.1), only β is relevant in evaluating the

proposal if she was to become a responder.) It turned out this part was not used for the analysis of this chapter. The roles of the subjects and the outcomes of the ultimatum game are only revealed at the end of the experiment after everyone made all the decisions during the experiment. We do not think this part has any influence on the choices in the later parts as will be discussed and analyzed in the following.

[4] How these bounds are derived has been discussed in Section 2.2. Again, it can be seen in the following example. Assume a proposer chooses A for Decision 2. This would imply for the proposer $u(95, 150) \geq u(100, 260)$ holds, which according to the utility form in equation (3.1) can be written as $95 - \alpha(150 - 95) \geq 100 - \alpha(260 - 100)$. It then further derives $\alpha \geq 0.05$. Doing the same for the other decisions gives the entries in the last column of Table 3.1.

Nr.	Option A	Option B	A is chosen, iff.
1	Yours:105; Other's: 150.	Yours: 100; Other's: 260.	$\alpha \geq -0.04$
2	Yours: 95; Other's: 150.	Yours: 100; Other's: 260.	$\alpha \geq 0.05$
3	Yours: 85; Other's: 150.	Yours: 100; Other's: 260.	$\alpha \geq 0.16$
4	Yours: 75; Other's: 150.	Yours: 100; Other's: 260.	$\alpha \geq 0.29$
5	Yours: 65; Other's: 150.	Yours: 100; Other's: 260.	$\alpha \geq 0.47$
6	Yours: 55; Other's: 150.	Yours: 100; Other's: 260.	$\alpha \geq 0.69$
7	Yours: 45; Other's: 150.	Yours: 100; Other's: 260.	$\alpha \geq 1.00$
8	Yours: 35; Other's: 150.	Yours: 100; Other's: 260.	$\alpha \geq 1.44$

Table 3.1: Menu 1. Envy (α) Level Measurement

Nr.	Option A	Option B	A is chosen, iff.
1	Yours: 85; Other's: 85.	Yours: 115; Other's: 15.	$\beta \geq 0.30$
2	Yours: 100; Other's: 100.	Yours: 115; Other's: 15.	$\beta \geq 0.15$

Table 3.2: Menu 2. Guilt (β) Level Measurement

utilities generated by Options A and B in Decisions 1 and 2. From a similar analysis as done for Menu 1, we can derive the necessary and sufficient condition on the proposer's β parameter for choosing A in each decision. These values are given in the last column, which is, again, not revealed to the subjects in the experiment.

In this part, subjects are informed before making their decisions in Menu 1 and 2, that in each pair of matched subjects, only one will be randomly selected to be the proposer, whose decision will be implemented. Others will be selected receivers, whose decisions will not influence any subjects' earnings. In addition, out of the 10 decisions in the two menus, only one decision will be randomly selected to be implemented at the end of the experiment. [5]

[5]It is also made clear to the subjects that the opponent they face for the decisions in Menu 1 are not the same as in Menu 2.

Stage 2: The Main Part of Games

	Give (G)	Take (T)
G	$100, 100$	$15, 115$
T	$115, 15$	$30, 30$

Game 2: The PD Game in the Experiment

For Part 2 of the experiment, the subjects are randomly matched to play the 2PD game. As described in Section 3.2.1, there is random rematching after the first round and we inform the subjects that their second-round opponents will be informed about their first-round choices. The two information mechanisms in the 2PD game, one revealing information about the opponent's type and the other not, are implemented in Treatment R2 and U2, respectively. In each of these two treatments, subjects play a 2PD game for 10 repetitions. The stage game played in each round of every repetition is Game 2.

Notice that Game 2 is a version of Game 1 with $c = 100$, $d = 115$, $f = 15$, and $s = 30$). With these specific parameters' values, condition (3.2) simplifies to

$$\beta \geq 0.15 \tag{3.11}$$

Next, recall that choosing A (B) twice in Menu 2 implies $\beta \geq 0.3$ ($\beta \leq 0.15$). Subjects who chose A twice in Menu 2 are labeled as 'Givers' in our experiment, and those who chose B twice as 'Takers'. According to the condition (3.11), For Givers are 'conditional cooperators' in Game 2 in the sense that for a pair of givers both giving is an equilibrium, while Takers are not. For the sake of part 2, we grouped all Givers and Takers in a 'Focus Pool', and subjects who chose A and B each exactly once in Menu 2,[6] in a 'Remain Pool'. Of course, subjects were not informed that they had been pooled in this way. For each repetition of the 2PD,

[6]There are two categories of these subjects in terms of their decisions in Menu 2. The first consists of those who chose B in Decision 1 but A in Decision 2. Such decisions cannot be rationalized in the IA model; their guilt level cannot be identified from the menu tests. The other category consists of subjects who chose A in Decision 1 and B in Decision 2, implying a $\beta \geq 0.15$, which also satisfies (3.2). To avoid noise in the inequity aversion measurement we do not label these subjects as Givers. This creates a 'safe guard' between the categories of Givers and Takers.

subjects are then randomly allocated to groups of six, with members of any such group coming from the same pool.[7] In each of the 10 repetitions of the 2PD game, subjects from each group of six are randomly rematched at the beginning of each round. In summary, at the beginning of part 2, subjects are allocated to either the focus pool or the remain pool, based on their decisions in part 1. At the start of each repetition they are allocated to groups of six, within their pool. In the two rounds of a repetition, they are twice randomly matched with another subject in the group of six. Subjects were only informed that they would be randomly rematched with others in the laboratory.

In the R2 treatment, when subjects are randomly paired with another member of the same group in each round of the 2PD, they are always informed about their opponents' types – Giver (who have chosen A twice in Menu 2) or Taker (who have chosen B twice in Menu 2), before making the decision on G or T in Game 2. In the other treatment U2, no information on the opponents' choices in Menu 2 (types) is revealed. In both treatments, subjects are informed at the beginning of the first round about the fractions of Givers and Takers in the group, which remains the same for each repetition of the 2PD. For the Remain Pool, the subjects are not labeled as 'Givers' or 'Takers'. Instead, they are randomly assigned a green label or a purple Label, irrespective of their decisions in Part 1. Every subject's label remains the same throughout Part 2. The subjects are informed about their paired opponents' label in the R2 treatment, but not in U2. In both treatments, a subject is always informed at the beginning of each round about how many subjects have a green label or a purple label in her group. Note that the labels defined in the Focus Pool (Givers and Takers) are meaningful, in the sense that they reflect a subject's past behavior, and moreover, their guilt levels. In contrast, the labels used in the Remain Pool (Green and Purple) do not have any real meanings. Thus, observations from the Remain Pool allow us to control for possible label effects on subjects' cooperativeness (i.e. whether facing an opponent with a same/different label makes one more/less cooperative). This control for a possible label effect in the Remain Pool allows us to isolate the influence of revealing opponents' true preference (revealed by their choices in Menu 2). In the following description of the design, we only explain the implementation for the Focus pool. The design applied to the Remain pool is exactly

[7] If the total number of the subjects in the focus pool is not a multiple of 6, the extra number of subjects from the Focus Pool are randomly selected and added to the Remain Pool.

the same as for the Focus pool, except that the type information on one being a Giver/Taker is replaced by the information on having a green/purple label.

The reason for informing subjects (in both the R2 and U2 treatments) how many of the other five subjects in the group are Givers/Takers is to make the fraction of each type in the population public information as assumed in our theory in Section 3.2, where the fraction of givers is denoted by ρ. We implement the strategy method in the first round of the 2PD, in the sense that every subject is asked to specify her decision on G/T for each possible composition of the population.[8] Immediately after entering the second round in each repetition, the subjects are informed about the true group composition in that repetition and asked to make only one decision.

In addition to the above two treatments, we implement two control treatments, denoted by R1 and U1, in which subjects play a one-round PD game (2) in 10 repetitions. As in Treatment R2 and U2, the subjects are assigned to a focus pool and remain pool according to their Stage 1 choices, and labeled as 'Givers' or 'Takers' (for the focus pool), or assigned a green or purple labels (for the remain pool). For each repetition of the PD game, the subjects from the same pool are randomly rematched. Control treatment U1 involves participation in a PD without information about the opponent's type and is used to test whether the classification of the subjects as Givers and Takers is consistent with their behavior in the PD game. In R1, subjects are informed about the opponent's type. This is used to test subjects' response to the type information without the forward looking strategic complications in the 2PD game.

For all treatments, it holds that in part 1, subjects are not informed about what is going to happen in part 2. Before entering part 2, each subject is informed that there is a chance that their decisions in the first part will be revealed to other subjects in the part to come. Each subject is given the option to either enter part 2 or to stay out.[9] We allow subjects to opt in or out of part 2 so that those concerned about having their previous choices revealed to others may keep them undisclosed by staying out of part 2.

[8]The potential combinations of the other five subjects in a group are (5/0), (4/1), (3/2), (2/3), (1/4), (0,5), with the two numbers in the parentheses denoting the number of Givers/Takers.

[9]We avoid letting the subjects know about the revelation of their decisions when making their decisions in part 1, because we want their choices to reveal true preferences rather than be influenced by strategic considerations related to future revelation.

The experiment was conducted at the CREED laboratory of University of Amsterdam in 2012 and 2013. It was implemented using the experimental software z-Tree (Fischbacher 2007). 220 subjects participated in 8 Sessions, with two sessions for each treatment R2, U2, R1, and U1. Each session lasted approximately one hour. The average earning per person was 12.82 Euro.

Out of our 220 subjects, 164 subjects chose either A twice or B twice in Menu 3.2, all of whom chose to enter Stage 2. Among the remaining 56 subjects, only one chose not to enter part 2. Due to the need of groups of size 6 (for R2 and U2), or size 2 (for R1 and U1) 152 (out of 164) subjects were randomly put in the focus pool. Of the remaining 68 subjects, 62 were put in the remain pool, and the other 6 subjects (due to the subject who chose not to enter Stage 2) did not participate in part 2 (but were required to stay in the laboratory until the session had finished.

Recall that only the composition of the focus pool (with only Givers and Takers) was intentionally designed to fit our theoretical set-up. In the following analysis, we therefore mainly focus on the behavior of the 152 subjects in the focus pool. The results presented in the next section therefore reflect observations in the focus group only, unless indicated otherwise.

3.4 Results

Table 3.3: Nr. of Givers and Takers in Each Treatment

	Giver		Taker		Total
	Obs	Percentage	Obs	Percentage	obs.
U1	20	55.56%	16	44.44%	36
R1	15	46.88%	17	53.13%	32
U2	20	47.62%	22	52.38%	42
R2	24	57.14%	18	42.86%	42

Across the four treatments, 79 subjects were classified as Givers and 73 as Takers, based on their Menu 3.2 choices. The numbers for each treatment are summarized in Table 3.3. In each treatment, there is roughly a 50-50 split between Givers and Takers. The hypothesis that the composition of Takers and Givers follows the same distribution across the four treatments is rejected neither by the Fisher exact test ($p = 0.746$) nor by the Pearson Chi-2 test ($p = 0.732$).

Before we use our laboratory observations to test the 'Ignorance is bliss' hypothesis we provide two validity checks to ensure that our experiment matches the theoretical set-up in Section 3.2: First, we investigate whether the classification of subjects as Givers or Takers is consistent with theory, in the sense that Givers are more cooperative than Takers in the prisoners' dilemma games. Second, we check whether the subjects in our experiment understood the information on types and responded to it in the way predicted. For these tests, we use the one round benchmarks U1 and R1.

Type Classification Consistent?

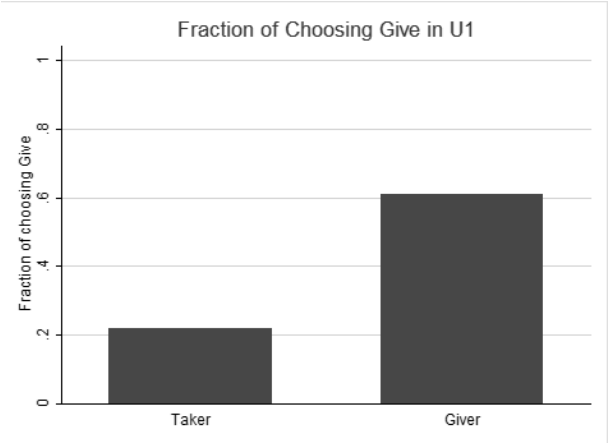

Figure 3.1: Fraction of Choosing 'Give' in U1

If correctly categorized, subjects labeled as Givers (based on their choices in Menu 3.2) will act as conditional cooperators in the one-shot Game 2, and subjects labeled as Takers will act as defectors. Figure 3.1) presents the fraction of each type that chose G in treatment U1, i.e., in a one-round PD without knowing the opponent's type. This shows that the Takers' cooperation rate is significantly lower than that of the Givers' (21.88% v.s. 61.00%, with $p < 0.01$ by Fisher exact test). This suggests that our classification is consistent with behavior expected from the types.[10]

[10]In contrast, subjects assigned with different labels in the remain group exhibit statistically indistinguishable cooperation rates in U1 (46.25 % v.s. 48.75%, $p = 0.87$).

Subjects' Reaction to Opponents' Types

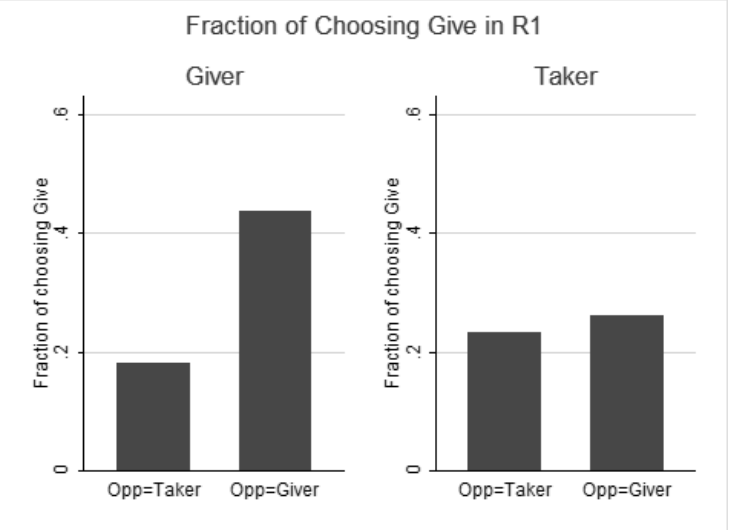

Figure 3.2: Fraction of Choosing 'Give' in R1

Next, we check whether subjects understood the type classification, by considering whether they reacted to the revealed type information in the way predicted. Following our game-theoretical predictions applied to treatment R1 (i.e., a one-round PD played after types have been revealed) Takers will not discriminate between the opponent's types, but instead will always choose the dominant strategy T against every opponent. In contrast, for Givers discrimination on opponents' types is predicted. They are expected to choose T when facing a Taker (anticipating that their opponents will choose T) but may choose G when the opponent is also a Giver in the hope of achieving the more efficient equilibrium (G, G). These predictions are supported by our observations from treatment R1 (as shown in Figure 3.2). This shows for Takers cooperation rates (fraction of G choices) of 23.17% against Takers and 26.14% against Givers, a statistically insignificant difference (Fisher exact test, $p = 0.72$). In contrast, while only 18.18% of the Givers' choices against Taker opponents were G, their frequency of choosing G against their own type was as high as 43.55%, which is more than twice as much. This discrimination by Givers on their opponents' types is statistically significant (Fisher exact test, $p < 0.01$). These re-

sults show that, subjects understood the relevance of the type information revealed to them and strategically responded to this information.[11]

The above two validity checks show that the classification of subjects in our experiment was consistent with their behavior and our communication of types to the subjects was well understood. We can therefore proceed to test whether as predicted by Proposition 1 ignorance is bliss in the 2PD, i.e., whether a sufficiently high fraction of Givers in the group yields higher cooperation rate in treatment U2 (no type information) than under R2 (types revealed).

Is Ignorance Bliss?

Table 3.4: Choices in Menu 3.1

last choice	Giver			Taker		
in A	Freq.	Percent	Cum.	Freq.	Percent	Cum.
0	40	50.63	50.63	8	10.96	10.96
1	16	20.25	70.89	51	69.86	80.82
2	3	3.80	74.68	1	1.37	82.19
3	3	3.80	78.48	3	4.11	86.3
4	4	5.06	83.54	3	4.11	90.41
5	1	1.27	84.81	4	5.48	95.89
6	1	1.27	86.08	0	0.00	95.89
7	2	2.53	88.61	1	1.37	97.26
8	9	11.39	100	2	2.74	100
Total	79	100		73	100	

Note: Last choice in A in Decision 0 means a subject never chose A in Menu 3.1.

Recall that the "Ignorance is bliss" prediction only holds when the fraction of Givers (ρ) in a group is above a certain threshold (cf. Proposition 1). The threshold is specified by the RHS of inequality (3.10) and depends on Givers' envy and guilt parameters, α^G and β^G.

To determine the envy parameter we use choices made in Menu 3.1. For each type, Table 3.4 displays for each decision in this menu, how many subjects made

[11] Neither the green nor purple label holders in the remain group showed any discrimination on their opponents' label information (for both label holders, the hypothesis that they respond two different label holders with the same cooperation rate is not rejected, with $p > 0.50$ for both.)

their last choice for Option A in Menu 3.1. Consider first the Takers. A majority of almost 70% of the takers chose Option A only in Decision 1, which implies an level of $\alpha \in [-0.04, 0.05)$. Note that the assumed value of the Takers' envy level in the theoretical model, $\alpha^T = 0$, lies exactly in this interval.

Relevant for the threshold for ρ is the envy parameter of Table 3.4 shows that 50% of the Givers never chose Option A (implying their $\alpha < -0.04$), and another 20% only chose A in Decision 1 (implying an $\alpha \in [-0.04, 0.05)$). To obtain a conservative estimate for the threshold we use the upper bound of these 70% Givers' elicited envy level, 0.05, as the estimate for α^G. For Givers' guilt level, we use the lower bound of all Givers' elicited guilt level, which is (by definition) 0.30. Because the threshold for ρ. as indicated by RHS of inequality (3.10), is weakly increasing in the value of α^G and decreasing in β^G, this choice of values may only overestimate the threshold, ensuring that the sample of cases (i.e., realized ρ values) used to test the "Ignorance is bliss" prediction fulfills the condition needed for the prediction to hold. Substituting $c = 100$, $d = 115$, $f = 15$, $s = 30$, $\alpha^G = 0.05$ and $\beta^G = 0.30$, inequality (3.10) yields $\rho \geq 0.57$. Given that our group size is 6, this means that for groups with 4 Givers, we should observe more cooperation under treatment U2 than under R2.[12]

For groups with four or more Givers in treatment U2 the PBE described in Lemma 2 predicts for Round 1 that both Givers and Takers choose G (to build a good reputation).In Round 2, Takers are predicted to choose T, and Givers choose G if and only if their opponents' first-round choice was G (which means that on the equilibrium path, Givers always choose G). In R2, in both rounds, Givers only cooperate with each other, and Takers always choose T. Therefore, in Round 1 both Takers and Givers are predicted to choose G more often under U2 than under R2. In Round 2, Givers are predicted to again play G more often under U2 than under R2, while Takers are predicted to choose G equally frequently (i.e., not at all) in the two treatments.

[12]If we used a slightly less conservative estimate, by taking $\alpha^G = 0$ and $\beta^G = 0.35$, the threshold derived by the RHS of inequality (3.10) would drop to 0.43, which would mean in groups with 3 Givers, higher cooperation rate could be reached under U2 than under R2. As in the groups with 4 Givers, however, higher cooperation rate in R2 than in U2 is also observed in these groups with 3 Givers.

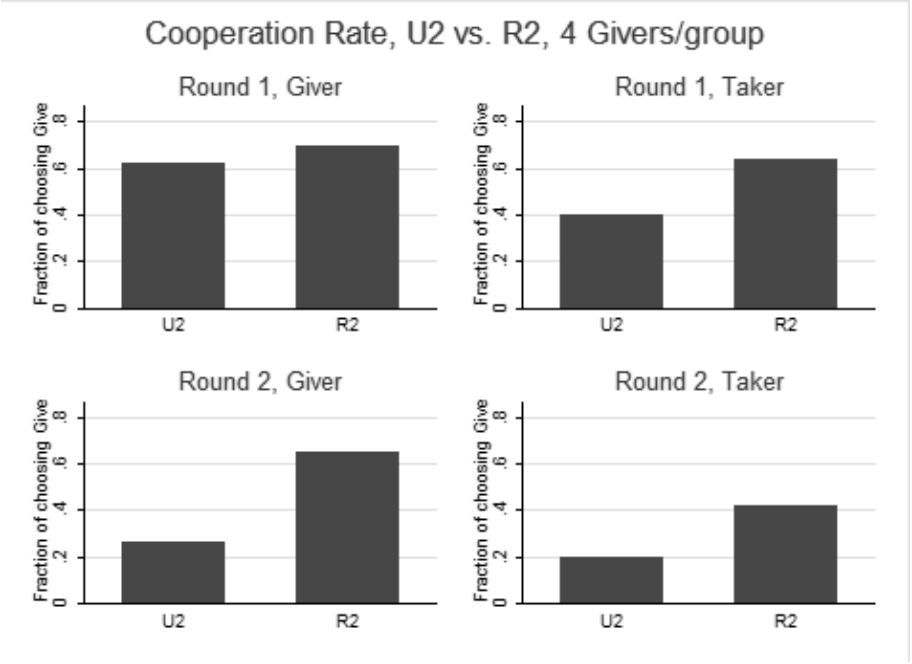

Figure 3.3: Cooperation Rate, U2 vs. R2. (4 Givers/group)

Our results are in stark contrast with these predictions. Figure 3.3 shows for all cases with 4 Givers in a group,[13] cooperation rates (fractions of G choices) decomposed by type and round. It shows for all cases higher cooperation under R2 than under U2. All differences between R2 and U2 are statistically significant (Fisher exact tests) with $p < 0.05$, except Givers' cooperation rates in Round 1, which are only marginally significant ($p = 0.07$). The effects are large. Both types exhibit drastic increases in cooperation rates when the type information is revealed. In round 2, Takers' cooperation rate doubled (20.00% vs. 41.67%), and Givers' cooperation rate even tripled (26.25% vs. 65.00%). These results are opposite to the "Ignorance is bliss" hypothesis predicted by the game-theoretical analysis.

[13]Recall that we used the strategy method in Round 1 with respect to group composition. Hence we had for all subjects a decision for the group composition of 4 Givers and 2 Takers. For Round 2, subjects only decided for the case of the actually realized group composition. We therefore, only use observations from the groups with 4 Givers and 2 Takers (there are 20 such groups under U2 and 30 under R2). Finally, in our experiment, under R2, there are no groups with 5 or more Givers, thus a comparison of the cooperation rates between R2 and U2 under such case of the group composition is unavailable.

To better understand the aspects of subjects' behavior that are causing this deviation from theory, we pool the data from the four sessions, and run two *probit* regressions to investigate how the Givers and Takers respond to different information in each round of the game. The two regressions are of the following forms, with eq (3.12) for the first-round behavior and eq (3.13) for the second-round behavior.

$$a_{i,1}^{*r} = \gamma_0^{T_i,t_i} + \gamma_1^{T_i,t_i} t_{-i1}^r I_{\{T_i=R1\}} + \gamma_2^{T_i,t_i} nG_i^r + \mu_i + \nu_i^r, \tag{3.12}$$

$$\begin{aligned} a_{i,2}^{*r} = {} & \delta_0^{T_i,t_i} + \delta_1^{T_i,t_i} t_{-i2}^r I_{\{T_i=R2\}} + \delta_2^{T_i,t_i} a_{-i2,1}^r + \delta_3^{T_i,t_i} a_{i,1}^r \\ & + \delta_4^{T_i,t_i} nG_i^r + \eta_i^{T_i,t_i} + \varepsilon_i^r, \end{aligned} \tag{3.13}$$

where $\nu_i^r, \varepsilon_i^r \overset{i.i.d}{\sim} \mathbb{N}(0,1)$. In the two regressions, t_{-i1}^r, t_{-i2}^r, $a_{-i2,1}^r$, and $a_{i,1}^r$ are four dummy variables, which respectively denote, in repetition r, whether subject i's round-1 opponent is a Giver, whether her round-2 opponent is a Giver, whether subject i's round-2 opponent chose G in round 1, and whether i herself chose G in round 1. nG_i^r denotes in repetition r, how many Givers there are in subject i's group. $I_{\{A\}}$ is the indicator function, which takes value 1 when event A is true, and 0 if otherwise. $\mu_i^{T_i,t_i} \overset{i.i.d}{\sim} \mathbb{N}(0, \sigma^2_{\mu^{T_i,t_i}})$ and $\eta_i^{T_i,t_i} \overset{i.i.d}{\sim} \mathbb{N}(0, \sigma^2_{\eta^{T_i,t_i}})$ are random effects for subject i in round 1 and round 2, respectively. The $a_{i,m}^{*r}$ are latent variables such that subject i chooses G in round m ($m \in \{1,2\}$) if and only if $a_{i,m}^{*r} > 0$. Each of the two regressions is run for each treatment $T_i \in \{R2, U2\}$ and each type of subjects $t_i \in \{Giver, Taker\}$.

The estimated marginal effects of each variable as well as their significance levels are shown in Table 3.5. From the game-theoretical analysis, in the first round only the Givers would discriminate between different opponent types (only when they know the opponent's types, i.e. in R2). The takers are expected to treat both types of opponents equally, by either always choosing G (in U2 when there are enough Givers in the group), or always choosing T (in R2 and in U2 when there are not enough Givers in the group). Hence, whether a Taker's opponent is a Giver is never predicted to have any effect on his cooperation rate in round 1. For both types, the number of Givers in the group is expected to have a positive effect in U2, because as the number increases beyond the threshold, the PBE switches from no player choosing G to everyone choosing G in the first round. In R2, having more Givers in a group increases the likelihood of a Giver being matched to another Giver, which means their cooperation rate is increasing in the number of Givers in the group. However, this effect does not exists for Takers when types are revealed,

because they are predicted to choose T anyway. The estimated marginal effects of the variables used for the first round in regression (3.12) are exactly as predicted: information that the opponent is a Giver (if available, i.e. in R2) has a highly significant ($p < 0.01$) positive marginal effect on Givers' cooperation rates, but not so for Takers. The number of Givers, as predicted, has a highly significant, positive marginal effect on the Givers' cooperation rates in both R2 and U2, and for Takers in U2 only.

In the second round, because the Takers are predicted to always choose T, none of the explanatory variables in regression (3.13) is theorized to have an impact on Takers' choices in R2. Givers in R2 are only predicted to respond to the opponents' type (and choose G only if the opponent is also a Giver), but not to any other information, including the opponent's or the own first round choices, or the number of the Givers in the group. In U2, theory predicts that Givers will not choose G if the number of Givers in the group is too low, and will switch to the strategies described in (3.4) where G might be chosen (depending on the opponents' and own choices in Round 1) otherwise. Hence, the number of Givers in the group is expected to have a positive impact on the Givers' cooperation rate. Also, because in strategy (3.4), a Giver chooses G only if the opponent and he both chose G in round 1, these two dummy variables are predicted to have a positive effect. The results show that the estimated marginal effects of the opponent's type information are again consistent with the predictions for both Givers and Takers. The predicted effects of the number of Givers are largely supported, though the positive effect expected for Givers in U2 is not found. For Givers in U2, the positive response to the own and opponent's giving behavior in round 1 is also as expected. The main deviation from the theoretical prediction is that this positive response is also observed for the other cases (Givers In R2; and Takers in both treatments), where it was not predicted. The effect of past choices is thus much stronger than predicted: in round 2, subjects seem to generally respond significantly positively, to their opponents' and their own round-1 choices in G.

Table 3.5: Probit Regression for Each Round in 2PD

	First Round							
	Giver				Taker			
	R2 (240)		U2 (200)		R2 (180)		U2 (220)	
	Pred.	M.E.	Pred.	M.E.	Pred.	M.E.	Pred.	M.E.
OppIsGiver	+	0.26***	- -	- -	0	0.10	- -	- -
nGivers	+	0.22***	+	0.13***	0	0.10	+	0.05***
	Second Round							
	Giver				Taker			
	R2 (240)		U2 (200)		R2 (180)		U2 (220)	
	Pred.	M.E.	Pred.	M.E.	Pred.	M.E.	Pred.	M.E.
OppIsGiver	+	0.27***	- -	- -	0	0.11	- -	- -
OppGave	0	0.40***	+	0.33***	0	0.32***	0	0.19**
I Gave	0	0.31***	+	0.28***	0	0.56***	0	0.28***
nGivers	0	0.04	+	0.01	0	-0.21	0	-0.02

Note: 1) The numbers of the observations for each regression are given in paranthesis after each treatment's name.
2) The two columns for each treatment give the game-theoretical sign prediction of the marginal effect for each variable and the estimated marginal effect.
3) ** and *** denote significance at the 5% and 1% level, respectively.

Recall that there is always a random rematching between the two rounds, so a subject's opponent's round-1 choice was very likely towards a different subject than herself. This suggests that indirect reciprocity may play a role in the observed strong reaction to opponents' previous choices. As for the strong positive effect in round 2 of the own first-round, this could be attributable to a strategic consideration. Knowing that my opponent would choose G with higher probability if I chose G in the previous round means that having chosen so makes it more likely that the (G,G) outcome is attainable, making a round-2 G choice more attractive to some.[14]

Given that we appear to observe indirect reciprocity in round 2 we consider whether this leads to strategic reputation building in round 1. If subjects have rational beliefs about others rewarding in round 2 a first-round cooperative choice G, straightforward backward induction implies a strategic response of more cooperation

[14]An alternative possibility is that the positive effect of the own previous choice reflects some internal consistency in subjects' choices. For example, there may be an intrinsic preference for cooperation that varies across people. As a consequence, subjects who are more likely to cooperate in round 1 are also more likely to cooperate in round 2. Note that this kind of preference is not captured by our type classification based on inequity aversion.

in round 1. This would imply that, compared to a one-shot game, subjects' should choose G more often in the first round of a 2PD. To test this hypothesis, we first pool the observations from R1 and the first round of R2, and those from U1 and the first round of U2, to two separate pools, R and U, respectively. In pool R, the opponent's type information is revealed, while it is not in U. Then, by introducing a variable $TwoR$, which indicates whether subject i is in a treatment with 2PD (i.e. $TwoR_i = 1$ for observations from R2 and U2, and 0 for observations from R1 and U1), we run the following regression on the pooled data:

$$a^{*r}_{i,1} = \lambda_0^{P_i,t_i} + \lambda_1^{P_i,t_i} t^r_{-i1} I_{\{P_i=R\}} + \lambda_2^{P_i,t_i} nG^r_i + \lambda_3^{P_i,t_i} TwoR + \zeta_i + \epsilon^r_i, \tag{3.14}$$

where $P_i \in \{U, R\}$ denotes the treatment pool for subject i. $\epsilon^r_i \overset{i.i.d}{\sim} \mathbb{N}(0,1)$. $\zeta_i \overset{i.i.d}{\sim} \mathbb{N}(0, \sigma^2_{zeta})$ is an individual random effect for subject i. The estimation is based on the assumption that Subject i chooses G (i.e. $a^r_{i,1} = 1$) if and only if $a^{*r}_{i,1} > 0$. If as we conjectured, the subjects behave strategically in the first round to build a reputation in order to be indirectly reciprocated by their future opponent, we should observe that whether there is a second round in the game has a positive influence on subjects' cooperation rates in round 1.

Table 3.6: Probit Regression (3.14), First-Round Behavior (1PD, 2PD pooled)

	First Round (1PD, 2PD pooled)							
	Giver				Taker			
	R (390)		U (400)		R (350)		U (380)	
	Pred.	M.E.	Pred.	M.E.	Pred.	M.E.	Pred.	M.E.
OppIsGiver	+	0.24***	- -	- -	0	0.03	- -	- -
nGiver	+	0.20***	+	0.10***	0	0.05	+	0.05***
TwoRound	+	0.03	+	-0.36	+	0.33	+	-0.02

Note: The prediction of the effect of $TwoR$ is based on a strategic anticipation of indirect reciprocity, as explained in the main text.

The estimated results are as shown in Table 3.6. The estimated effects of the opponent's type and the number of Givers are similar to what we obtained using observations from R2 and U2 only, and again, consistent with the game-theoretical prediction. Interestingly, the marginal effect of there being a second round is not found to be significant, which means we do not have any evidence showing that the subjects strategically anticipate indirect reciprocity.

The observation that the subjects do not strategically respond to others' indirect reciprocity, even if it takes only one step of backward-induction, reminds us of the fact that the way that people make decisions does not always follow the process assumed in a game-theoretical analysis. An alternative approach to understanding human decision making is used in the biological literature, where strategies are defined and investigated with respect to their performance under evolutionary pressures. Such strategies may be backward looking, adaptive or a combination of both. One model that is relevant to the setup in our experiments is the so-called image-scoring model, proposed by Nowak and Sigmund (1998).

Adapted to our game, this model assumes that each subject i evaluates an image score (s_{-it}) of her round-t opponent. This image score summarizes what i observes about this opponent, including the type and previous choice (whenever available). An image-scoring strategy then prescribes choosing G if and only if the opponent's image score exceeds a given threshold τ^G or τ^T, dependent on the opponent's type. These thresholds define the strategy. For a decomposed PD, Nowak and Sigmund (1998) show that image-scoring strategies evolve to a stable, perfectly discriminating cases where indirect reciprocity is the norm and (only) people who have been cooperative to others in the past are responded to cooperatively. To start, any subject has a default image score of 0 in her opponent's eyes. Otherwise, if she is revealed to be a Giver, the image score is increased by 1, and if revealed to be a Taker, the image score is decreased by 1. The previous choice, which is always revealed to an opponent, also affects a subject's image score in round 2: her score is increased by k after choosing G, and decreased by k after choosing T, where $k \in (0, +\infty)$ measures how much the subjects weigh their opponents' previous choices compared to their types.

With these image-score-updating rules, any opponent starting the first round in U1 or U2 , has an image score of 0, and any opponent starting round 1 in $R1$ or $R2$ has an image score of either 1 or -1, depending on whether she is a Giver or a Taker. At the beginning of round 2, a subject's image score can be k or $-k$ in U2, or a value in $\{-1-k, -1+k, 1-k, 1+k\}$ in R2.

Using our laboratory observations, we find out that, for any value of k, the best fitting values of τ^G and τ^T are such that τ^G is below 0 and above all possible negative image scores, while the Takers' threshold τ^T is larger than $1+k$ (which is the largest possible image score for any subject). Hence, our data yield a model where Givers choose G when facing any opponent with a non-negative image score

while Takers never choose G. This model predicts 61% of the Givers' decisions and 72% of the Takers' decisions correctly. Note that the "always T" strategy for Takers is consistent with their type description in the game-theoretical perspective. The Givers' strategy can be summarized as "I will cooperate with you as long as I have no evidence of you being non-cooperative". This strategy is more generous than the game-theoretical interpretation of the Givers which is "I will cooperate with you if and only if you also cooperate with me". The difference between the two strategies is that, a Giver adopting the former strategy would always cooperate with a completely stranger, whose choice history being unknown, however, another Giver who adopts the latter "conditional cooperation" strategy would not choose cooperate unless she expects her opponent to choose G with a sufficiently high probability.

3.5 Summary

This chapter presented a game-theoretical model to analyze the effects of revealing players' type information in a repeated social dilemma game. The PBE predicts that, when there are enough cooperative types in the population, it is better (in terms of efficiency) to not reveal the players' types. This theoretical 'Ignorance is Bliss' seems to provide support to practices outside of the laboratory, such as the expungement of juvenile criminal records. However, our laboratory experiment suggests the opposite. When revealing type information, efficiency is boosted, because all types cooperate more in a repeated prisoner's dilemma game.

We provide evidence on several behavioral factors, which we suspect may have driven subjects' behavior from the prediction. First and foremost, subjects appear to apply indirectly reciprocal strategies. Second, this indirect reciprocity is not met with strategic reputation building, which seemingly implies strong bounds on our subject' rationality. This is because reputation building would have required only one step of strategic thinking.

Indirect reciprocity and bounds on the rationality of decision makers are often elements of evolutionary models. In such models, strategies are viewed as having evolved in a survival-of-the-fittest environment. This can lead to seemingly non-profitable strategies (at least in the short run), and decisions following simple rules. We were inspired by such findings, in particular related to the image-scoring model of indirect reciprocity and fit this model to our data. This model turns out to successfully describe choices by a majority of both types of our subjects.

3.6 Appendix

3.6.1 Proof of Lemma 2

Proof. We need to show that for the strategies specified in eq (3.7), (3.8), neither type of player has an incentive to deviate. (For convenience in the proof, we denote by c', d', f', s' the utility for a Giver when the outcome in the 1PD is $(A_1, A_2) = (G, G), (T, G), (G, T), (T, T)$, respectively, where A_1 and A_2 are, respectively, the choices of the Giver and her opponent. These values can easily be calculated using c,d,f,s,α^G and β^G.)

In round 2, Takers will always choose T regardless of information and other players' decisions, because T is their dominant action. They therefore will not deviate from the specified round-2 strategy.

For a Giver if either her or her opponent's first round choice is not G, her opponent (regardless of being a Giver or a Taker) will choose T in round 2. The best response is to also choose T. Instead, when a Giver and her opponent both chose G in round 1, and given that both types choose G in round 1, the rational belief of the opponent being a Giver is ρ. her opponent, in this situation, will choose G if she is a Giver, and T if she is a Taker. Then the expected utility for the Giver from choosing G is $\rho c' + (1-\rho)f'$, and the expected utility from choosing T is $\rho d' + (1-\rho)s'$. Hence, if the following incentive compatibility condition holds, Givers do not have an incentive to deviate in Round 2.

$$\rho c' + (1-\rho)f' \geq \rho d' + (1-\rho)s' \tag{3.15}$$

Substituting $c' = c$, $f' = f - \alpha^G(d-f)$, $d' = d - \beta^G(d-f)$ and $s' = s$, the above condition reduces to

$$\rho \geq \frac{s - f + \alpha^G(d-f)}{c + s - d - f + (\alpha^G + \beta^G)(d-f)} \tag{3.16}$$

In round 1, given that everyone chooses G, a Giver's best response is G. So the Givers do not have any incentive to deviate from the given strategy.

For a Taker, if she chooses G in round 1, her round-2 opponent will choose G if it is a Giver, and T if it's a Taker, which means she would end up receiving d with probability ρ and s with probability $1-\rho$. If she chooses T in round 1, her round-2 opponent will always choose T, and her earning from round 2 will be s for sure. So the incentive compatibility condition for the Taker to not deviate from their

first-round strategy is:

$$\rho d + (1 - \rho)s \geq s \tag{3.17}$$

which simplifies to

$$\rho \geq \frac{d - c}{d - s} \tag{3.18}$$

Thus, if conditions (3.16) and (3.18) hold, i.e. $\rho \geq \max\{\frac{d-c}{d-s}, \frac{s-f+\alpha^G(d-f)}{c+s-d-f+(\alpha^G+\beta^G)(d-f)}\}$, the strategies specified in eq (3.7) and (3.8) form a PBE of the 2PD under the type-unrevealing information mechanism. □

3.6.2 Instructions (FocusGroup,TR2)

Part 1

In this experiment, all the payoffs will be displayed in points. The exchange rate is 100 points=2 Euro. You are now in Part 1 of the experiment.

In this part, you will be randomly paired to another participant. One of each pair will be randomly selected to be the proposer, while the other will become the responder.

The proposer has to make a decision on how to split a total payoff of 100 points between the proposer and the responder.

If the responder accepts the proposal, the two participants' payoffs from Part 1 will be determined as specified in the proposal. If the responder rejects, both the proposer and the responder will receive 0 points from Part 1.

Participants will only learn their roles at the very end of the whole experiment. So each of you needs to make your decision both as a proposer and as a responder. This means that you need, first, to specify your proposal as if you are selected to be the proposer; and second, to decide as if you are a responder.

When deciding as a responder, you will not know what an offer you will receive from your proposer, so you need to tell us which offers you would reject, and which you would accept. To do this you can choose a threshold. If in the proposal from your paired proposer, he/she has offered you a number of points that is not lower than the threshold chosen by you, the proposal will be accepted, and you will get the amount as offered in the proposal and your paired proposer will get 100-your points. However, if the amount your opponent proposes to offer you is below your threshold, the proposal will be rejected, and each of you will earn 0 points.

When deciding as a proposer, you need to specify two numbers, one is the amount of points for your paired responder, the other for yourself (the two must add up to 100). If it turns out that the amount you proposer to offer your responder is equal or higher than his/her threshold, your proposer is accepted, and your responder and you will respectively receive the amounts as proposed by you. However, if the amount you propose to offer your responder is lower than his/her threshold, your proposal will be rejected, and each of you will receive 0 points.

At the end of the experiment, your role (proposer or responder) will be randomly assigned and your decision for your true role will be used to determine you and your

paired participant's payoffs in Part 1.

Part 2 In this part of the experiment, you will again be randomly paired with another participant in this room. The participant paired with you is most likely a different one from the one with whom you were paired with in Part 1.

In Part 2, you will be asked to make 10 decisions in total. Each decision involves a choice between an Option A and an Option B, taking the form of:

Option A	**Option B**
Your Payoff:	Your Payoff: .....
Other's Payoff:	Other's Payoff: .....

The options refer to payment to you and one of the other participants in this experiment. For each option, two amounts will be displayed: one amount that you will receive yourself, and one amount that the "Other", the participant you are paired with, will receive.

You will be given two menus, Menu 1 with 8 decisions and Menu 2 with 8 decisions. In Menu 1 you will be randomly paired with a different participant than the one you are paired with for Menu 2. Your decision in this part will determine your payoff and the payoff of the other participant you are paired with in the following way.

Within each pair, one participant will be randomly selected with a probability of 50%, to be the Proposer, and the other will be the Receiver. If you are chosen to be the proposer, one out of the in total 10 decisions you have made in Part 2 will be randomly selected with equal probability. You will receive the amount you decided to receive in that chosen decision and your paired receiver will receive the amount you decided to give to "Other". Your paired receiver's decision in Part 2 will have no influence at all. ' If you are chosen to be the receiver, one out of the 10 decisions made by your paired proposer in Part 2 will be randomly selected with equal probability. Your paired proposer will receive the amount he/she decided to receive in that chosen decision and you will receive the amount he/she decided to give to "Other". In this case, Your decisions in Part 2 will have no influence at all. Your decision will not influence in any way what role you will be assigned to and which decision will be implemented for payment.

Example:

Option A	**Option B**
Your Payoff: 20	Your Payoff: 30
Other's Payoff: 40	Other's Payoff: 50

In the given example, suppose your choice is Option B and your paired participant's choice is A. Then if you are randomly selected to be the proposer, the payoffs in your pair will be determined by your choice: you will receive 30 and your paired participant (the receiver) will receive 50, which is the amount you chose to give "other".

If your paired participant is randomly selected to be the proposer, his/her choice for the selected decision will determine the payoffs for your pair: he/she will receive 20 and you will receive 40. In either case, the receiver's decision in Part 2 will not have any influence.

Before Part3

You have finished the first two parts of the experiment. There is a Part 3, which is the LAST part of this experiment. In Part 3, you may still earn a substantial amount of money.

During Part 3, some of your choices made in the previous parts may be shown to the other participants. If so, this will be always done anonymously.

You may choose whether or not to enter the next part. If you choose not to enter, your earning from Part 3 will be 0. If you do enter Part 3, you may add to the amount you have earned in the previous two parts. You cannot lose money in Part 3.

The results of the previous parts will only be shown at the very end of the whole experiment, and everyone will only be paid then.

Part 3

You are now in Part 3 of the experiment. In this part, you will play a 2-round game. The 2-round game will be repeated for 10 times. Only one out of the 10 repetitions of the 2-round game will be randomly selected to be paid for Part 3. Please read these instructions carefully.

2-Round Game The game has two rounds. In each round, you will be randomly paired with another participant in this room, and you two will play a game.

How are you paired with others? For each repetition, you will be firstly assigned in to a group consisting of 6 participants (you and five others). In each round, you will be randomly paired with one of the other five participants in your group. The probability of being paired with any of them will be equal. Note that the random matching will be done at the beginning of each round, so your Round-1 paired participant is most likely different from your Round-2 paired participant. After each repetition of the 2-round game finishes, participants will be randomly assigned into different groups of 6 again. So in different repetitions of the 2-round game, you will be probably with different 5 other group members.

What is the game in each round? In each round you will play a game, which will be the same for both rounds, with your paired participant of that round. Both of you have two choices, Give or Take, and you two will make your choices simultaneously. Your payoff and that of your paired participant will be determined according to the following rules.

	If your PP's choice is Give	If your PP's choice is Take
If your choice is Give	Your payoff: 100; Your PP's payoff: 100.	Your payoff: 15; Your PP's payoff: 115.
If your choice is Take	Your payoff: 15; Your PP's payoff: 115.	Your payoff: 30; Your PP's payoff: 30.

Choice for Round 1 will be Revealed in Round 2 Your decision in Round 1 will be revealed to your Round-2 paired participant at the beginning of Round 2, before the decisions in that round are made. In a same way, your paired participant's decision in Round 1 will also be revealed to his/her Round-2 paired participant then.

Information Box: There will be an information box on the screen when you make your decisions. You will find some information about you and/or your paired opponent in that box.

(The following part for the Focus Group only)

In Round 1 of the 2-round game, you are randomly matched with another participant in your group.

The grouping is based on decisions made in Menu 2 (see below), in Part 2 of today's experiment. You were asked to make two decisions between Option A and B in Menu 2.

Givers & Takers: All six participants (including yourself) in your group chose either Option A in both decisions or Option B in both decisions. In the following, we will refer to the participants that chose A twice in Menu 2 as **Givers** for short, and that chose B twice as **Takers**. The intuition behind the names is that, choosing A instead of B in Menu 2 menus to give up part of ones own earnings to increase others' income, and choosing B means to take away some of other's earnings to increase one's own earning.

Because you will be randomly rematched in Round 2 with one of the other five participants again, you may want to know how many of those five others are givers and ho many are takers. For this reason, we will let you choose (between Give and Take) separately for every possible combination of Giver and Takers. (Your decision for the case corresponding to the true group composition will be implemented and revealed to your Round-2 paired participant). Of course, you don't have to make different choices for different group compositions.

Giver or Taker? In **every round**, you can observe whether your paired participant is a Giver or a Taker in the information box. In Round 2, apart from your current paired participant's Round-1 choice, you can also observe whether he/she was paired with a Giver or a Taker in Round 1 in the information box.