Reciprocity and trust: 
Personality psychology meets behavioral economics

Ricardo Andrés Guzmán · Rodrigo Harrison · 
Nureya Abarca · Mauricio G. Villena

Received: date / Accepted: date

Abstract We propose a model of strong reciprocity and trust that combines elements of personality psychology and behavioral economics. In the model, positive reciprocity and negative reciprocity are components of individual utility functions, while trust is an individual bias that distorts the beliefs about the trustworthiness of others. Underlying personality traits determine the functional form of the utility functions and the magnitude of the biases.

We tested the model by means of an economic experiment, using 212 college students as experimental subjects. First, each subject completed a psychometric questionnaire to measure his levels of positive reciprocity, negative reciprocity, and trust. Following this, the subjects played 65 rounds of a sequential prisoner’s dilemma with random re-matching and payoffs that changed from round to round. From the game behavior of each subject we estimated his utility function and the bias in his beliefs. The experimental results support the hypotheses of the model; the psychometric measures of reciprocity determined the individual utility functions, while the psychometric measures of trust determined the individual biases.

Keywords Strong reciprocity · trust · personality · psychometrics · revealed preferences. 
JEL codes: C72, C92, D03.

Ricardo Andrés Guzmán, corresponding author. 
Centro de Investigación en Complejidad Social (CICS), Facultad de Gobierno, UDD. Address: Avenida Plaza 680, San Carlos de Apoquindo, Las Condes, Santiago, Chile. Tel.: +562 327-97-01 E-mail: rguzman@udd.cl

Rodrigo Harrison 
Instituto de Economía, Pontificia Universidad Católica de Chile. Address: Avenida Vicuña Mackenna 4860, Macul, Santiago, Chile. E-mail: harrison@uc.cl

Nureya Abarca 
Escuela de Administración, Pontificia Universidad Católica de Chile. Address: Avenida Vicuña Mackenna 4860, Macul, Santiago, Chile. E-mail: nabarca@faceapuc.cl

Mauricio G. Villena 
Escuela de Negocios, Universidad Adolfo Ibáñez. Address: Diagonal Las Torres 2640, Peñalolén, Santiago, Chile. E-mail: mauricio.villena@uai.cl
Personality psychology and behavioral economics share an interest in strong reciprocity and trust (Fehr & Gächter 2000; Perugini et al., 2002; Sobel 2005). Each discipline has approached these phenomena using its proprietary set of models and empirical methods. Their relative advantages and disadvantages suggest that a synthesis between both approaches could be useful to further our understanding of strong reciprocity and trust (Almlund et al. 2011; Caplan, 2003; Ferguson et al. 2011; Heckman 2011; Heckman et al. 2008). Personality psychology could contribute its ability to describe what economists call preferences and beliefs in terms of a reduced set of personality traits. These traits could be measured using psychometric techniques, which have shown high predictive power in a wide range of real-life behaviors, including reciprocal and trusting behaviors. Behavioral economics, in turn, could contribute a theoretical framework that accounts for the situation (“the game”) and material incentives. This framework consists of utility functions that allow for heterogeneity in reciprocal preferences and probabilistic models of individual beliefs, plus game theoretic models representing social interactions in which reciprocity and trust play a role; for example, the trust game, the ultimatum game, and the prisoner’s dilemma.

Along these lines, we propose a synthetic model of reciprocity and trust that connects elements of personality psychology and behavioral economics. The model has four parts: (1) a sequential prisoner’s dilemma that represents a scenario of reciprocal interaction (Clark & Sefton 2001); (2) a utility function with reciprocal preferences that depend on personality traits (Charness & Rabin 2002); (3) a model of subjective (probabilistic) beliefs about how likely other players are to cooperate whose bias also depends on personality traits, and (4) a quantal response function (McKelvey & Palfrey 1998) describing how players choose among three alternative strategies, to cooperate conditionally, to cooperate unconditionally, and to defect unconditionally. In other words, this model aims to capture the way in which personality manifests itself given different incentives in the context of a prisoner’s dilemma. The model also aims to explain reciprocal preferences and beliefs of trustworthiness as functions of personality traits.

We also report an empirical test of the synthetic model. This test combines empirical methods of both disciplines: (1) a psychometric questionnaire specially designed to measure positive reciprocity, negative reciprocity, and trust (Dohmen et al. 2008), and (2) an economic experiment, in which the subjects played a sequential prisoner’s dilemma with random re-matching and payoffs that changed from round to round. From the game behavior of each experimental subject we estimated his utility function and the bias in his probabilisitic beliefs. The experimental results support the hypotheses of the model: the psychometric measures of reciprocity determined the individual utility functions, while the psychometric measures of trust determined the individual biases.

The rest of this paper proceeds as follows. In Section 2 we briefly compare the approaches of personality psychology and behavioral economics to the study of strong reciprocity and trust. In Section 3 we develop the synthetic model. In Section 4 we present the materials and methods of
the empirical test. In Section 5 we report the results of the study. Finally, in Section 6 we offer some conclusions.

2 Two alternative approaches to strong reciprocity and trust

In personality psychology, positive reciprocity, negative reciprocity, and trust are modeled as personality traits, or as combinations of higher-order personality traits (Dohmen et al. 2008; Perugini et al. 2003). Psychometricians typically measure personality traits using self-report questionnaires. They also look for statistical relations between personality traits and behaviors and life outcomes. Recent studies have found that measured levels of positive and negative reciprocity predict various behaviors in the workplace, earned incomes, and probability of employment (Dohmen et al. 2009; Dur et al. 2010; Raymond et al. 2012). Trust as a personality trait, on the other hand, has been widely studied (Jones et al. 1997). It has predictive power in many aspects of life, including individual economic behaviors and performance (Butler et al. 2009).

Personality traits predict a wide range of behaviors and life outcomes, averaged over many situations and occasions (Funder, 2008; Roberts 2009). They are, however, bad predictors of behaviors and outcomes in specific situations. This is because personality models do not take into account the behavioral effects of context and material incentives (Funder 2008). Recently, psychologists have begun to study how personality manifests itself in different situations (e.g., Fleeson 2007; Fournier et al. 2008), but not how personality is modulated by with economic incentives.

More fundamentally, personality models have been criticized for having a tenuous theoretical basis (Almlund et al. 2011; Blanton & Jaccard 2006). Also, psychometrics faces severe identification problems: it can detect correlations between personality traits and behaviors or outcomes, but it often fails to establish causality (Borghans et al. 2011; Almlund et al. 2011; Heckman 2011). Furthermore, self-report questionnaires are vulnerable to misrepresentation: examinees can manipulate their answers to present a desirable image, or they may disclose a distorted self-image (Viswesvaran & Ones 1999).

In behavioral economics, positive and negative reciprocity are modeled as social preferences (Cox et al. 2007; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher 2006; Rabin 1993). Trust, on the other hand, is modelled as an individual probabilistic belief about the trustworthiness of others (Ashraf et al. 2006; Buchan et al. 2008; Eckel & Wilson 2004; Fetchenhauer & Dunning 2009). To measure social preferences and trust, researchers conduct experiments in which people play games for money, such as the trust game, the ultimatum game, and the prisoner’s dilemma (Camerer & Fehr 2004). Some researchers use this experimental data to classify individual utility functions into types: selfish, altruistic, inequity-averse, positively reciprocal, negatively reciprocal, etc. (e.g., Burlando & Guala 2005; Charness & Rabin 2002; Engelmann & Strobel 2004; Fischbacher et al. 2001; Kurzban & Houser 2001; Rodríguez-Sickert et al. 2008). Other re-
searchers use experimental data to estimate the complete functional form of the utility functions (e.g., Andreoni & Miller 2002; Andreoni et al. 2009; Charness & Rabin 2002; Fisman et al. 2005; Goeree et al. 2002).

Compared to personality psychology, behavioral economics has a firmer theoretical basis, which is provided by choice theory and game theory (Camerer et al. 2003, part two). Unlike personality models, choice theory and game theory take into account the behavioral effects of context (“the game”) and material incentives. Also, economic experiments can identify causality by means of controlled variation (Falk & Heckman 2009), while the double-anonymous experimental design plus the use of material incentives discourage the misrepresentation of preferences and beliefs (Bardsley et al. 2010, Chapter 6 and p. 232).

Nevertheless, behavioral economics has a serious weakness: many of its experimental findings lack external validity. This means that the experimental findings do not extend to “real life”, at least not in the direct, unconditional way demanded by some critics of experimental economics. The lack of external validity is particularly prevalent in experiments measuring social preferences (Levitt & List 2007); there are, however, some positive findings. For example, a recent study by Barr and Serneels (2009) found a bicausal relation between worker productivity and reciprocating behavior in the investment game.

In comparison to experimental economics, psychometrics has a much better record in predicting real-life behavior. To our knowledge, only one study has done the comparison directly. Anderson et al. (2011) compared the predictive power of psychometric measures of personality and experimental estimates of risk-aversion and delay-discounting. The authors found that, in combination, personality measures outperformed estimated economic preferences in predicting credit scores, job persistence, driving accidents, body-mass index, and smoking habits.

The predictive power of personality traits is not limited to real-life behaviors: personality traits also predict behavior in many economic experiments (Brocklebank et al. 2011). These experiments include the dictator game (Ben-Ner et al. 2004; Ben-Ner & Kramer 2011), the trust game (Evans & Revelle 2008; Ben-Ner & Halldorsson 2010; Burks et al. 2004; Gunnthorsdottir et al. 2002), the ultimatum game (Brandstätter & Königstein 2001), the prisoner’s dilemma (Boone et al. 1999, 2002; Hirsh & Peterson 2009; Pothis et al. 2011), and the public good game (Kurzban & Houser 2001; Perugini et al. 2010). There are few negative findings (Swope et al. 2008).

3 A model of strong reciprocity and trust

The model has four parts, which we describe in the following sections: (1) a sequential prisoner’s dilemma that represents a scenario of reciprocal interaction (Section 3.1); (2) a utility function with reciprocal preferences that depend on personality traits (Section 3.2); (3) subjective probabilistic beliefs about how likely other players are to cooperate whose bias also depends on personality traits (Section 3.3), and (4) a quantal response function (Section 3.4) describing
how players choose among three alternative strategies, to cooperate conditionally, to cooperate unconditionally, and to defect unconditionally.

3.1 A sequential prisoner’s dilemma

Two players participate in a sequential prisoner’s dilemma (Clark & Sefton, 2001). The players can perform two actions: cooperate or defect. One player moves first and decides, blindly, whether to cooperate or not. Before knowing this decision, the second mover chooses his response among three alternative strategies:

1. Cooperate unconditionally, regardless of whether the first mover has cooperated or defected.
2. Cooperate conditionally, if and only if the first mover has cooperated.
3. Defect unconditionally, regardless of whether the first mover has cooperated or defected.

The extensive form of the game is presented in Figure 1. Letters “c” and “d” denote the players’ actions, and function $\pi(\cdot)$ represents the money payoffs of the game. The payoffs are symmetric: if the first mover performs action $a_1$, and the second mover responds with action $a_2$, the first mover obtains $\pi(a_1, a_2)$ while the second mover obtains $\pi(a_2, a_1)$. The money payoffs satisfy the following inequalities, which imply that the game is a prisoner’s dilemma:

$$\pi(d, c) > \pi(c, c) > \pi(d, d) > \pi(c, d) \geq 0.$$ 

If the players maximized expected income, the game would have a unique Nash equilibrium in strictly dominant strategies: the first mover defects blindly and the second mover defects unconditionally.

3.2 A utility function with reciprocal preferences

We focus our attention on the second mover, because he has the option to reward or punish the actions of the first mover; that is, to reciprocate the behavior of the first mover. The second mover’s utility function is given by

$$u(\pi_2, \pi_1) = \begin{cases} 
(1 - r^+)\pi_2 + r^+\pi_1 & \text{if the 1st mover cooperated blindly}, \\
(1 + r^-)\pi_2 - r^-\pi_1 & \text{if the 1st mover defected blindly}. 
\end{cases}$$ \hspace{1cm} (1)

This is an adapted version Charness and Rabin’s utility function (Charness & Rabin 2002). In the formula, $\pi_1$ and $\pi_2$ are the payoffs obtained by the first and the second mover, respectively. Parameter $r^+$ represents the second mover’s positive reciprocal preference, while $r^-$ represents his negative reciprocal preference. In principle, we let $r^+, r^- \in \mathbb{R}$. The second mover’s utility will be strictly increasing in his own payoff if and only if $r^+ < 1$ and $r^- > -1$. 

Fig. 1: The sequential prisoner’s dilemma. The players can perform two actions: cooperate (c) or defect (d). The payoffs are symmetric: if the first mover performs action $a_1$, and the second mover responds with action $a_2$, the first mover obtains $\pi(a_1, a_2)$ while the second mover obtains $\pi(a_2, a_1)$. The payoffs satisfy $\pi(d, c) > \pi(c, c) > \pi(d, d) > \pi(c, d) \geq 0$.

The utility function given in equation (1) is flexible enough to capture many intuitive preference structures. Six archetypal cases are illustrative:

1. If $r^+ = r^- = 0$, the second mover is selfish:
   \[
   u(\pi_2, \pi_1) = \pi_2.
   \]

2. If $r^+ \in (0, 1]$ and $r^- = -r^+$, the second mover is an altruist:
   \[
   u(\pi_2, \pi_1) = (1 - r^+)\pi_2 + r^+\pi_1.
   \]

3. If $r^- > 0$ and $r^- = -r^+$, the second mover is spiteful:
   \[
   u(\pi_2, \pi_1) = (1 + r^-)\pi_2 - r^-\pi_1.
   \]

4. If $r^+ \in (0, 1]$ and $r^- = 0$, the second mover is a pure positive reciprocator:
   \[
   u(\pi_2, \pi_1) = \begin{cases} 
   (1 - r^+)\pi_2 + r^+\pi_1 & \text{if the 1st mover cooperated blindly}, \\
   \pi_2 & \text{if the 1st mover defected blindly}.
   \end{cases}
   \]
5. If \( r^- > 0 \) and \( r^+ = 0 \), the second mover is a pure negative reciprocator.

\[
\begin{align*}
\mu(\pi_2, \pi_1) &= \begin{cases} 
\pi_2 & \text{if the 1st mover cooperated blindly,} \\
(1 + r^-)\pi_2 - r^-\pi_1 & \text{if the 1st mover defected blindly.}
\end{cases}
\end{align*}
\]

6. If \( r^+ \in (0, 1] \) and \( r^- > 0 \), the second mover is simultaneously a positive and a negative reciprocator [see equation (1)].

The second mover’s reciprocal preferences depend on two personality traits that represent his underlying levels of positive and negative reciprocity. Let \( R^+ \) be his positive reciprocity trait, and let \( R^- \) be his negative reciprocity trait, where \( R^+, R^- \in \mathbb{R} \). Both traits are random variables and may or may not be correlated. Assume that \( \mathbb{E}[R^+] = \mathbb{E}[R^-] = 0 \), and \( \text{var}[R^+] = \text{var}[R^-] = 1 \).

Reciprocal preferences relate to reciprocity personality traits in the following way:

\[
\begin{align*}
r^+ &= \rho^+_0 + \rho^+_1 R^+ + \rho^+_2 R^- = \mathbf{R}^T \mathbf{\rho}^+, \quad (2) \\
r^- &= \rho^-_0 + \rho^-_1 R^+ + \rho^-_2 R^- = \mathbf{R}^T \mathbf{\rho}^-., \quad (3)
\end{align*}
\]

where \( \mathbf{\rho}^+, \mathbf{\rho}^- \in \mathbb{R}^3 \) are vectors of parameters common to all members of the population. Since \( \mathbb{E}[R^+] = \mathbb{E}[R^-] = 0 \), parameters \( \rho^+_0 \) and \( \rho^-_0 \) are the reciprocal preferences of the average second mover.

We formulate the following hypotheses regarding the relation between reciprocal preferences and reciprocity traits.

**Hypothesis 1** The average second mover’s utility function is strictly increasing in his payoff:

\[
\begin{align*}
\rho^+_0 &< 1, \quad (4) \\
\rho^-_0 &> -1. \quad (5)
\end{align*}
\]

**Hypothesis 2** The average second mover is a positive and negative reciprocator:

\[
\begin{align*}
\rho^+_0 &\in (0, 1], \quad (6) \\
\rho^-_0 &> 0. \quad (7)
\end{align*}
\]
Hypothesis 3 The positive reciprocal preference is increasing in the positive reciprocity trait, and does not depend on the negative reciprocity trait:

\[ \rho_1^+ > 0, \]
\[ \rho_2^+ = 0. \]

Hypothesis 4 The negative reciprocal preference is increasing in the negative reciprocity trait, and does not depend on the positive reciprocity trait:

\[ \rho_1^- > 0, \]
\[ \rho_1^- = 0. \]

3.3 Subjective beliefs

Let \( p \in [0,1] \) be the true probability that the first mover cooperates blindly. The second mover believes this probability is \( \mu \), which is given by

\[ \mu = \frac{p \exp(t)}{1 - p + p \exp(t)}, \]

where \( t \in \mathbb{R} \) is the second mover’s trust coefficient. By construction, \( \mu \in [0,1] \) and \( \mu \) is an increasing function of \( t \). This means that trustful second movers have large values of \( t \), while distrustful second movers have low values of \( t \).

Three cases are worth noting:

1. If \( t = 0 \), the second mover guesses the true value of \( p \); that is, \( \mu = p \).
2. If \( t > 0 \), the second mover overestimates \( p \); that is, \( \mu > p \).
3. If \( t < 0 \), the second mover underestimates \( p \); that is, \( \mu < p \).

The second mover’s trust coefficient depends on a personality trait that represents his underlying level of trust. Let \( T \) be his trust trait, where \( T \in \mathbb{R} \). The trust trait is a random variable, and may or may not be correlated to the reciprocity traits. Assume that \( \text{E}(T) = 0 \), and \( \text{var}(T) = 1 \).

The trust coefficient relates to the trust trait in the following way:

\[ t = \tau_0 + \tau_1 T = T^T \tau, \]

where \( \tau \in \mathbb{R}^2 \) is a vector of parameters common to all members of the population. Since \( \text{E}(T) = 0 \), parameter \( \tau_0 \) is the trust coefficient of the average second mover.

We formulate the following hypotheses regarding the relation between the trust coefficient and the trust trait.

Hypotheses 5 The trust coefficient is increasing in the trust trait:

\[ \tau_1 > 0. \]
3.4 The quantal response function

The second mover chooses his strategy non-deterministically, skewing the probabilities toward the strategies that provide higher expected utility. Let $S = \{CU, CC, DU\}$ be the set of available strategies, where CU means “cooperate unconditionally,” CC means “cooperate conditionally,” and DU means “defect unconditionally.” Denote by $E[u(s)]$ the subjective expected utility from choosing strategy $s$, and by $q(s)$ the probability of choosing strategy $s$. This probability is given by a quantal response function:

$$q(s) = \frac{\exp(\lambda E[u(s)])}{\sum_{x \in S} \exp(\lambda E[u(x)])},$$  \hspace{1cm} (15)$$

where $\lambda > 0$ is a parameter common to all members of the population.

By construction, $q(s)$ is increasing in $E[u(s)]$. Also, the larger $\lambda$, the more likely the second mover will choose the strategy that maximizes his expected utility. In the limiting case in which $\lambda$ tends to infinity, the second mover acts as an expected utility maximizer. In contrast, if $\lambda$ equals zero, he chooses all strategies with equal probability. For these reasons, we call $\lambda$ the rationality parameter.

To calculate the expected utility of the different strategies, we combine the payoffs in Figure 1, the utility function defined in equation (1), the subjective probability defined in equation (12), and the definitions of $r^+$, $r^-$, and $t$ given in equations (2), (3), and (13):

$$E[u(CU)] = \mu \pi(c, c) + (1 - \mu) \left[ (1 + R^T \rho^-) \pi(c, d) - R^T \rho^- \pi(d, c) \right]$$ \hspace{1cm} (16)$$

$$E[u(CC)] = \mu \pi(c, c) + (1 - \mu) \pi(d, d)$$ \hspace{1cm} (17)$$

$$E[u(DU)] = \mu \left[ (1 - R^T \rho^+) \pi(d, c) + R^T \rho^+ \pi(c, d) \right] + (1 - \mu) \pi(d, d),$$ \hspace{1cm} (18)$$

where

$$\mu = \frac{p \exp(T^T \tau)}{1 - p + p \exp(T^T \tau)}. \hspace{1cm} (19)$$

Together, equations (15)–(19) constitute a model of strategic behavior in which the probability of the second mover choosing a particular strategy is a function of his personality traits and the payoffs of the game. The values of the parameters ($\rho^+$, $\rho^-$, $\tau$, and $\lambda$) are to be determined empirically.

4 Materials and methods

4.1 Subjects and procedure

The experimental sessions took place at the Pontifical Catholic University of Chile between May and June 2010. A total of 212 college students from various academic majors volunteered as subjects. They were 18.9 years old on average, with a standard deviation of 1.9. Ninety-two
of the subjects were female, and all subjects were native Spanish speakers. We conducted 10 experimental sessions, which were attended by between 16 and 24 subjects each. The sessions were carried out in a computer room equipped with z-Tree, a program for economic experiments (Fischbacher 2007). We used a double-anonymous experimental design to mitigate the effects of the social desirability bias, and random re-matching to eliminate the incentives for reputation building (Bardsley et al. 2010). No communication was allowed between the subjects during the experiment.

Each experimental session was divided into two stages. At the beginning of each stage, the session coordinator distributed printed instructions which he then read aloud (the original instructions written in Spanish, plus an English translation, are available in an online appendix). After reading the instructions, the coordinator answered questions from the subjects, and begun the first stage.

In the first stage of the experiment the subjects completed a psychometric questionnaire specifically designed to measure reciprocity and trust. The reciprocity/trust psychometric questionnaire was taken from the 2005 wave of the German Socio-Economic Panel (Dohmen et al. 2008). This questionnaire includes nine statements: three refer to positive reciprocity, three refer to negative reciprocity, and three refer to trust. Table 1 displays the nine statements in the order presented to the subjects. Since all subjects were native Spanish speakers, we translated the questionnaire into Spanish (the translation is available in an online appendix). Each subject was asked to rate, on a 5-point Likert scale, his degree of agreement with each statement. Additionally, the subjects completed the NEO-PI-R personality inventory (Costa & McCrae 1985), whose results we do not analyze in this paper. After finishing the first stage, the subjects took a five-minute break.

During the second stage of the experiment, the subjects played 65 rounds of a sequential prisoner’s dilemma: 5 practice rounds, numbered −5 to −1, followed by 60 rounds for real money, numbered 1 to 60. The payoffs of the game changed from round to round, as shown in Table 2.

At the beginning of each round, the subjects were randomly matched. Since the number of rounds exceeded the number of subjects, two subjects could play together more than once during the experiment. Anonymity prevented the subjects from knowing when this happened.

In each round, each subject had to make two decisions: (1) what to do if he was given the role of first mover, and (2) what to do if he was given the role of second mover. Recall that the options for a first mover are to cooperate or defect blindly, and the options for a second mover are to cooperate unconditionally, to cooperate conditionally, or to defect unconditionally. The subjects had to make their decisions in private and before knowing the role that they would play in the current round.

Once both members of a pair made their decisions, z-Tree “flipped a coin” and assigned the roles of first and second mover. The subjects’ payoffs were computed accordingly. At this point, each subject was informed of the role he was given, whether his partner cooperated or not, and their respective payoffs. The subject was also informed of his accumulated earnings so far.
Table 1: The reciprocity/trust psychometric questionnaire

<table>
<thead>
<tr>
<th>Item</th>
<th>Statement</th>
<th>Trait</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>In general, one can trust people.</td>
<td>Trust</td>
</tr>
<tr>
<td>2</td>
<td>If someone does me a favor, I am prepared to return it.</td>
<td>Positive reciprocity</td>
</tr>
<tr>
<td>3</td>
<td>If I suffer a serious wrong, I will take revenge as soon as possible, no matter what the cost.</td>
<td>Negative reciprocity</td>
</tr>
<tr>
<td>4</td>
<td>These days you cannot rely on anybody else.</td>
<td>Trust</td>
</tr>
<tr>
<td>5</td>
<td>I am ready to undergo personal costs to help somebody who helped me before.</td>
<td>Positive reciprocity</td>
</tr>
<tr>
<td>6</td>
<td>If somebody puts me in a difficult position, I will do the same to him/her.</td>
<td>Negative reciprocity</td>
</tr>
<tr>
<td>7</td>
<td>When dealing with strangers it is better to be careful before you trust them.</td>
<td>Trust</td>
</tr>
<tr>
<td>8</td>
<td>I go out of my way to help some body who has been kind to me before.</td>
<td>Positive reciprocity</td>
</tr>
<tr>
<td>9</td>
<td>If somebody insults me, I will insult him/her back.</td>
<td>Negative reciprocity</td>
</tr>
</tbody>
</table>

Note: Items 2 and 3 are reverse coded. The third column was not presented to the subjects.

The subjects were paid in private upon leaving the session. No show-up fee was offered to the subjects. Total game earnings ranged between CLP 10,000 (≈ USD 20) and CLP 20,000 (≈ USD 40), approximately (CLP is the acronym for Chilean pesos).

4.2 Method of estimation

We estimated the model using the maximum likelihood method.

Let $y_{ij}(s) = 1$ if subject $i$ chose strategy $s$ in round $j$, and $y_{ij}(s) = 0$ if he did not. The log-likelihood function is defined as follows:

$$
L(\beta) = \sum_{i=1}^{212} \sum_{j=1}^{60} \sum_{s \in S} y_{ij}(s) \ln q_{ij}(s, \beta),
$$

where $q_{ij}(s, \beta)$ probability that subject $i$ chooses strategy $s$ in round $j$, $S = \{UC, CC, DU\}$ is the set of available strategies, and $\beta$ is a vector of parameters. This vector of parameters is given by

$$
\beta = \begin{bmatrix}
\rho^+ \\
\rho^- \\
\lambda \\
\tau
\end{bmatrix},
$$

and it must be determined empirically by maximizing $L(\beta)$. 
<table>
<thead>
<tr>
<th>Round</th>
<th>$\pi(d,d)$</th>
<th>$\pi(d,c)$</th>
<th>$\pi(c,d)$</th>
<th>$\pi(c,c)$</th>
<th>Round</th>
<th>$\pi(d,d)$</th>
<th>$\pi(d,c)$</th>
<th>$\pi(c,d)$</th>
<th>$\pi(c,c)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>−5</td>
<td>100</td>
<td>350</td>
<td>0</td>
<td>150</td>
<td>31</td>
<td>100</td>
<td>250</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>−4</td>
<td>200</td>
<td>300</td>
<td>0</td>
<td>250</td>
<td>32</td>
<td>50</td>
<td>350</td>
<td>0</td>
<td>150</td>
</tr>
<tr>
<td>−3</td>
<td>300</td>
<td>450</td>
<td>0</td>
<td>350</td>
<td>33</td>
<td>150</td>
<td>450</td>
<td>0</td>
<td>400</td>
</tr>
<tr>
<td>−2</td>
<td>100</td>
<td>300</td>
<td>0</td>
<td>200</td>
<td>34</td>
<td>100</td>
<td>450</td>
<td>0</td>
<td>400</td>
</tr>
<tr>
<td>−1</td>
<td>250</td>
<td>400</td>
<td>0</td>
<td>300</td>
<td>35</td>
<td>50</td>
<td>250</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>1</td>
<td>250</td>
<td>350</td>
<td>0</td>
<td>300</td>
<td>36</td>
<td>100</td>
<td>350</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>300</td>
<td>400</td>
<td>0</td>
<td>350</td>
<td>37</td>
<td>150</td>
<td>450</td>
<td>0</td>
<td>300</td>
</tr>
<tr>
<td>3</td>
<td>150</td>
<td>450</td>
<td>0</td>
<td>350</td>
<td>38</td>
<td>150</td>
<td>400</td>
<td>0</td>
<td>250</td>
</tr>
<tr>
<td>4</td>
<td>150</td>
<td>400</td>
<td>0</td>
<td>200</td>
<td>39</td>
<td>200</td>
<td>400</td>
<td>0</td>
<td>250</td>
</tr>
<tr>
<td>5</td>
<td>150</td>
<td>350</td>
<td>0</td>
<td>200</td>
<td>40</td>
<td>100</td>
<td>300</td>
<td>0</td>
<td>150</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>300</td>
<td>0</td>
<td>250</td>
<td>41</td>
<td>50</td>
<td>450</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>300</td>
<td>0</td>
<td>150</td>
<td>42</td>
<td>50</td>
<td>300</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>8</td>
<td>150</td>
<td>350</td>
<td>0</td>
<td>300</td>
<td>43</td>
<td>200</td>
<td>400</td>
<td>0</td>
<td>350</td>
</tr>
<tr>
<td>9</td>
<td>100</td>
<td>450</td>
<td>0</td>
<td>350</td>
<td>44</td>
<td>200</td>
<td>350</td>
<td>0</td>
<td>250</td>
</tr>
<tr>
<td>10</td>
<td>250</td>
<td>400</td>
<td>0</td>
<td>350</td>
<td>45</td>
<td>50</td>
<td>350</td>
<td>0</td>
<td>250</td>
</tr>
<tr>
<td>11</td>
<td>50</td>
<td>200</td>
<td>0</td>
<td>150</td>
<td>46</td>
<td>100</td>
<td>400</td>
<td>0</td>
<td>300</td>
</tr>
<tr>
<td>12</td>
<td>150</td>
<td>450</td>
<td>0</td>
<td>250</td>
<td>47</td>
<td>200</td>
<td>450</td>
<td>0</td>
<td>300</td>
</tr>
<tr>
<td>13</td>
<td>150</td>
<td>250</td>
<td>0</td>
<td>200</td>
<td>48</td>
<td>100</td>
<td>350</td>
<td>0</td>
<td>300</td>
</tr>
<tr>
<td>14</td>
<td>50</td>
<td>350</td>
<td>0</td>
<td>100</td>
<td>49</td>
<td>50</td>
<td>450</td>
<td>0</td>
<td>250</td>
</tr>
<tr>
<td>15</td>
<td>100</td>
<td>350</td>
<td>0</td>
<td>250</td>
<td>50</td>
<td>100</td>
<td>400</td>
<td>0</td>
<td>250</td>
</tr>
<tr>
<td>16</td>
<td>250</td>
<td>450</td>
<td>0</td>
<td>350</td>
<td>51</td>
<td>150</td>
<td>400</td>
<td>0</td>
<td>300</td>
</tr>
<tr>
<td>17</td>
<td>250</td>
<td>450</td>
<td>0</td>
<td>300</td>
<td>52</td>
<td>100</td>
<td>400</td>
<td>0</td>
<td>150</td>
</tr>
<tr>
<td>18</td>
<td>350</td>
<td>450</td>
<td>0</td>
<td>400</td>
<td>53</td>
<td>100</td>
<td>450</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>19</td>
<td>100</td>
<td>400</td>
<td>0</td>
<td>200</td>
<td>54</td>
<td>50</td>
<td>450</td>
<td>0</td>
<td>150</td>
</tr>
<tr>
<td>20</td>
<td>150</td>
<td>400</td>
<td>0</td>
<td>350</td>
<td>55</td>
<td>50</td>
<td>250</td>
<td>0</td>
<td>150</td>
</tr>
<tr>
<td>21</td>
<td>50</td>
<td>400</td>
<td>0</td>
<td>250</td>
<td>56</td>
<td>100</td>
<td>350</td>
<td>0</td>
<td>150</td>
</tr>
<tr>
<td>22</td>
<td>150</td>
<td>300</td>
<td>0</td>
<td>200</td>
<td>57</td>
<td>200</td>
<td>300</td>
<td>0</td>
<td>250</td>
</tr>
<tr>
<td>23</td>
<td>150</td>
<td>350</td>
<td>0</td>
<td>250</td>
<td>58</td>
<td>300</td>
<td>450</td>
<td>0</td>
<td>350</td>
</tr>
<tr>
<td>24</td>
<td>50</td>
<td>400</td>
<td>0</td>
<td>350</td>
<td>59</td>
<td>100</td>
<td>300</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>25</td>
<td>300</td>
<td>450</td>
<td>0</td>
<td>400</td>
<td>60</td>
<td>250</td>
<td>400</td>
<td>0</td>
<td>300</td>
</tr>
<tr>
<td>26</td>
<td>150</td>
<td>450</td>
<td>0</td>
<td>200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>50</td>
<td>450</td>
<td>0</td>
<td>400</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>150</td>
<td>300</td>
<td>0</td>
<td>250</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>50</td>
<td>350</td>
<td>0</td>
<td>200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>50</td>
<td>300</td>
<td>0</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Payoffs in Chilean pesos (CLP). Rounds −5 to −1 are practice only.
As of May 1st, 2010, 1 CLP = 0.0019 USD.
The probability that subject $i$ chooses strategy $s$ in round $j$ is given by the following expression:

$$q_{ij}(s) = \frac{\exp(\lambda E[u_{ij}(s, \beta)])}{\sum_{x \in S} \exp(\lambda E[u_{ij}(x, \beta)])}, \quad (21)$$

where

$$E[u_{ij}(s, \beta)] = \begin{cases} 
\mu_{ij}(\tau)\pi_j(c, c) + (1 - \mu_{ij}(\tau))\left[(1 + R_i^T \rho^-)\pi_j(c, d) - R_i^T \rho^- \pi_j(d, c)\right] & \text{if } s = \text{UC}, \\
\mu_{ij}(\tau)\pi_j(c, c) + (1 - \mu_{ij}(\tau))\pi_j(d, d) & \text{if } s = \text{CC}, \\
\mu_{ij}(\tau)\left[(1 - R_i^T \rho^+)\pi_j(d, c) + R_i^T \rho^+ \pi_j(c, d)\right] + (1 - \mu_{ij}(\tau))\pi_j(d, d) & \text{if } s = \text{DU}, 
\end{cases} \quad (22)$$

and

$$\mu_{ij}(\tau) = \frac{p_j \exp(T_i^T \tau)}{1 - p_j + p_j \exp(T_i^T \tau)}. \quad (23)$$

The above equations are instances of equations (15)–(19). Variables $R_i$ and $T_i$ are subject $i$’s vectors of personality traits, as measured by the reciprocity/trust psychometric questionnaire. These scores are standardized, so that they have zero mean and unitary variance. Variable $p_j$ is the empirical probability that the first mover cooperates blindly in round $j$. We define $p_j$ as follows:

$$p_j = \frac{M_j}{N} \quad (24)$$

Where $N$ is the number of subjects that participated in the experiment, and $M_j$ is the number of subjects that chose to cooperate blindly in round $j$.

5 Results

5.1 Exploratory analysis

The empirical distributions of personality traits, as measured by the reciprocity/trust psychometric questionnaire, are displayed in Figure 2.¹ These distributions closely resemble those reported by Dohmen et al. (2008). Most subjects obtained high scores in the positive reciprocity trait and low scores in the negative reciprocity trait. The distribution of the trust trait is more symmetric but slightly inclined to the left. Table 3 shows the correlations between the three scores. Positive and negative reciprocity do not correlate with each other, while trust correlates positively with positive reciprocity, and negatively with negative reciprocity.

The dynamics of cooperation in the sequential prisoner’s dilemma is summarized in Figure 3. The figure shows the fraction of subjects who chose a particular strategy by round of play.

¹ The database with the results of the psychometric test and the prisoner’s dilemma experiment can be downloaded from here: https://mega.co.nz/#!1ZVUyRSS!UUhYhUHrOLT5l9va2hmZshQ_qYKRUhDHltQ49H1nrU
Fig. 2: Empirical distribution of personality traits among the subjects.

Table 3: Correlations matrix of personality traits

<table>
<thead>
<tr>
<th></th>
<th>Positive reciprocity</th>
<th>Negative reciprocity</th>
<th>Trust</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive reciprocity</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative reciprocity</td>
<td>0.07</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Trust</td>
<td>0.16*</td>
<td>-0.36*</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Significant at the 5% level.

Table 4: Overall distribution of strategies

<table>
<thead>
<tr>
<th>Role</th>
<th>Strategy</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st mover</td>
<td>Cooperate blindly</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>Defect blindly</td>
<td>0.69</td>
</tr>
<tr>
<td>2nd mover</td>
<td>Cooperate unconditionally</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Cooperate conditionally</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>Defect unconditionally</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Observe that all types of cooperative strategies decline throughout the game. This is a typical result in cooperative dilemma experiments (Ledyard 1995). Averaging over subjects and rounds, we get the overall distribution of strategies. This is shown in Table 4.

What effects do incentives have on the second movers’ strategy choices? We specify incentives as follows:

\[
\text{temptation} = \pi_{dc} - \pi_{cc}, \\
\text{risk} = \pi_{dd} - \pi_{cd}, \\
\text{reward} = \pi_{cc} - \pi_{dd}.
\]

_Temptation_ is what the second mover gains by defecting rather than cooperating when the first mover cooperates. _Risk_ is what the second mover loses by cooperating rather than defecting when the first mover defects. _Reward_ is what both players gain by cooperating together rather than defecting together. Figure 4 shows the effects of incentives on the second movers’ strategy choices. Observe that a higher temptation reduces unconditional and conditional cooperation,
and increases unconditional defection. Risk, on the other hand, reduces only unconditional cooperation, albeit weakly. Finally, reward increases unconditional and conditional cooperation, and reduces unconditional defection.

Figure 5 shows the effects of personality on the behavior of second movers. As can be seen in the figure, positive reciprocity, negative reciprocity, and trust have a strong effect on the second movers’ strategic choices.

5.2 Estimation of the model and hypotheses testing

Table 5 presents the maximum-likelihood estimates of the parameters of the model. Replacing the estimated values into equations (2), (3), and (13) we can express the reciprocal preferences

---

2 To calculate the standard errors, p-values, and confidence intervals we used a semi-parametric bootstrapping method. Each iteration consisted of three steps. First, we resampled the 212 subjects, with replacement (a subject is a tuple of three elements: a positive reciprocity trait, a negative reciprocity trait, and a trust trait). Second, we
and the trust coefficient as functions of the reciprocity traits:

\[ r^+ = 0.27 + 0.06R^+ - 0.12R^-, \]

\[ r^- = 1.05 + 0.15R^-, \]

\[ t = 0.20T. \]

These equations only include the values statistically significant at the 10% level. The results confirm hypotheses 1, 2, 4, and 5, while hypothesis 3 is only partially confirmed (equality (9) is falsified by the data).
In summary:

1. The average second mover’s utility is strictly increasing in his payoff ($\rho^+ > 0$, $\rho^- < -1$)
2. The average second mover is a positive reciprocator ($\rho^+_0 \in (0, 1]$) and negative reciprocator ($\rho^-_0 > 0$)
3. The positive reciprocal preference is increasing in the positive reciprocity trait ($\rho^+_1 > 0$), and decreasing the negative reciprocity trait ($\rho^-_1 < 0$).
4. The negative reciprocal preference is increasing in the negative reciprocity trait ($\rho^-_2 > 0$), and does not depend on the positive reciprocity trait ($\rho^+_2 = 0$).
5. The trust coefficient is increasing in the trust trait ($\tau_1 > 0$).
### Table 5: Maximum-likelihood estimates of the parameters of the model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Coeff.</th>
<th>Std. err.</th>
<th>p-value</th>
<th>[90% conf. inter.]</th>
<th>[95% conf. inter.]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pos. reciprocal preference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant ((\rho_0^+))</td>
<td>0.27</td>
<td>0.02</td>
<td>0.00</td>
<td>0.24</td>
<td>0.30</td>
</tr>
<tr>
<td>Pos. reciprocity trait ((\rho_1^+))</td>
<td>0.06</td>
<td>0.02</td>
<td>0.00</td>
<td>0.04</td>
<td>0.09</td>
</tr>
<tr>
<td>Neg. reciprocity trait ((\rho_2^+))</td>
<td>-0.12</td>
<td>0.03</td>
<td>0.00</td>
<td>-0.17</td>
<td>-0.08</td>
</tr>
<tr>
<td>Neg. reciprocal preference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant ((\rho_0^-))</td>
<td>1.05</td>
<td>0.26</td>
<td>0.00</td>
<td>0.73</td>
<td>1.57</td>
</tr>
<tr>
<td>Pos. reciprocity trait ((\rho_1^-))</td>
<td>0.04</td>
<td>0.04</td>
<td>0.22</td>
<td>-0.02</td>
<td>0.11</td>
</tr>
<tr>
<td>Neg. reciprocity trait ((\rho_2^-))</td>
<td>0.15</td>
<td>0.10</td>
<td>0.10</td>
<td>0.00</td>
<td>0.32</td>
</tr>
<tr>
<td>Trust coefficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant ((\tau_0))</td>
<td>0.09</td>
<td>0.16</td>
<td>0.60</td>
<td>-0.18</td>
<td>0.36</td>
</tr>
<tr>
<td>Trust trait ((\tau_1))</td>
<td>0.20</td>
<td>0.08</td>
<td>0.01</td>
<td>0.06</td>
<td>0.34</td>
</tr>
<tr>
<td>Rationality parameter ((\lambda))</td>
<td>0.0109</td>
<td>0.0014</td>
<td>0.00</td>
<td>0.0089</td>
<td>0.0132</td>
</tr>
</tbody>
</table>

Note: Bootstrapped standard errors, p-values, and confidence intervals (10,000 draws).

### 6 Conclusion

Even though the domain of our synthetic model is limited to the current experimental design, the empirical validity of the model is relevant to both behavioral economics and personality psychology. From the point of view of behavioral economics, the synthetic model is a small step toward a theory of heterogeneous preferences and beliefs, which should be explained (at least in part) as a result of personality differences. This is important, because a theory of heterogeneous preferences and beliefs is an essential ingredient of any full-fledged model of human behavior. From the point of view of personality psychology, the empirical validity of the synthetic model indicates that economic theory can be used to understand how the behavioral manifestations of personality are affected by the situation; specifically, by material incentives. Ultimately, the results of this study suggest that a synthesis between behavioral economics and personality psychology is possible and potentially fruitful.

### Acknowledgments

For their valuable comments and suggestions, we thank Leda Cosmides, James C. Cox, Giuseppe Attanasi, Aaron Lukaszewski, and Carlos Rodríguez-Sickert. María Ignacia Rivera and Eric González provided research assistance. This research was conducted with the financial support of Fondecyt grant #1120387, and Proyecto Anillo SOC1101.
References


