Intuition versus Reflection in a Cooperation Dilemma
Evidence for value-based choice decisions

Juana Castro
Supervisor: Salvador Soto Faraco

1. Introduction

Classical economic models study human behavior under the assumption that any action is motivated by the expected utility it will convey to the subject. This concept of utility refers to the “amount” of satisfaction or benefit a subject can obtain by taking any action (deciding to purchase something, investing or saving money or helping a friend). In other words, the utility accounts for the relative value assigned to an action that serves also as a measure of individual preferences; how much a subject values a piece of chocolate, swimming in a pool, or an airplane ticket to Thailand in 3 years. Nonetheless, given the difficulty of measuring utility or satisfaction, the classical economics approach has mainly focused in monetary variables for which an objective value can be accounted to study decision-making.

How would a person decide whether to help or not a stranger? Classical economics would argue this behavior to be irrational unless the benefit from helping others would exceed the cost of providing help. Given countless examples of this kind of “irrational” behaviors in the real life and an evident impossibility of explaining them with the classical economics approach, in the late 90’s, a non-orthodox economic approach to study decision-making started developing. In this field the study of human decision-making shifted from the study of only market-driven incentives (quantifiable variables) to include in the analysis other variables that can better explain social behavior using an experimental approach. New economic modeling is then now used in social experiments to study social preferences for cooperation, power, justice, fairness, etc.; all social constructions that influence decision-making processes.
Recently, a special interest for speed in decision-making has emerged. A new paradigm for economic decision-making materialized with Kahneman’s (2011) work on the dual process theory that suggests that intuitive and reflexive behaviors compete at a cognitive level, to guide individual decision-making. This theory proposes intuition (also referred to as ‘System I’) as a fast and automatic process, and reflection (‘System II’) as a slow and calculated process in which an individual tries to consider all possible outcomes before reaching a decision.

Under this paradigm, recent literature has tried to establish whether cooperation in social dilemmas is an intuitive or a reflective behavior by measuring RTs (reaction times) for different choice options. For this matter, social dilemmas have been modeled as economic decisions in which responses corresponding to selfish or cooperative behavior are measured in different kinds of economic games such as the Dictator, Prisoners Dilemma, Ultimatum and Public Goods Games (Rubinstein, 2007; Piovesan & Wengström, 2009; Cappelletti et al., 2011; Grimm et al., 2011; Rubinstein, 2011; Rand et al., 2012; Fiedler et al., 2013; Zaki & Mitchell, 2013; Lotito et al., 2013; Cappelen et al., 2014; Lohse et al., 2014; Rand et al., 2015; Fehr et al., 2015, Rand et al., 2015).

The interest in measuring RTs is that it is a quantifiable behavioral measure that could be used as a proxy for intuitive/reflective cognitive processes. Under the dual process theory paradigm faster RTs, compared with slower ones, would allow the differentiation between intuitive and reflective behaviors in value-based choice decisions.

However, research on this field has produced somehow inconsistent results; some studies claiming that cooperative decisions are faster (and hence more intuitive) than selfish ones (Zaki & Mitchell, 2013; Lotito et al., 2013; Rand et al., 2012; Cappelen et al., 2014;) and some reporting results in the opposite direction (Piovesan & Wengström, 2009; Fiedler et al., 2013; Lohse et al., 2014).

This contradiction in the obtained results is explained in Fehr and colleagues (2015) by the oversight of two important sources of variability in the analysis when one just considers the correlation of RTs with individual choice. The first oversight is the variability in choice problems given different incentives: depending on the payment structure a selfish/cooperative choice can become more/less attractive and faster/slower than the other. Fehr et al. explored this effect using a discrete Dictator and Public Goods Game for which they manipulated the monetary incentives. The authors did this manipulation by repeatedly varying the difference (in monetary gains) between choosing the cooperative option or the selfish one, making the decision to cooperate/not cooperate more/less attractive. In their experimental design Fehr et al. used three different levels of incentives (high, medium, low) and compared RTs for each level of incentives. Results showed that RTs are faster as the utility difference between two choices is bigger, and slower when the two choices involve similar monetary outcomes. This was interpreted as the existence of a steeper competition between choices when the outcomes are closer.

The second source of variability investigated in Fehr et al.’s experimental design was the so-
called strength of preference\(^1\) that refers to the individual differences in preference (one person might prefer to cooperate, whereas some might adopt the strategy of never cooperating), something that was not controlled in prior studies. After controlling in each case for the individual strength of the preference, they found no evidence of one type of choice (cooperative or selfish) being overall faster than the other. From these results, Fehr et al. finally conclude that RTs cannot be used to infer different cognitive processes (intuitive/reflexive) for decision-making.

Rand et al (2015) add evidence to support the conclusion that RTs are an inadequate proxy for intuitive versus reflective cognitive processes. This evidence is based on various studies using a continuum value based choice design for cooperation decisions where they assert that RTs are mostly driven by decision conflict (related to the results found by Rubinstein, 2007). The authors found responses in the extremes of the continuum (pure selfish and cooperative decisions) to be the fastest and explain choosing the middle options as result of cognitive conflict; individuals that reported more conflict took longer choosing their response and were less likely to choose an extreme response. Additionally in a study inducing time pressure and time delay, they find that choosing responses at the extreme points of the continuum was strongly related to the time pressure condition, and found an overall significant negative relationship between choosing extreme responses and RTs.

As we have attempted to illustrate above, results from studies measuring RTs for cooperation dilemmas suggest that speed of response is driven by similarity of choice-options, strength of preference, and cognitive conflict (Fehr et al. 2015 and Rand et al. 2015 correspondingly). This last evidence finds that the originally observed differences in RTs for cooperative and selfish choices disappear after controlling for above mentioned variables and they use this evidence to claim RTs may not be a good measure to disentangle cognitive processes (intuitive and reflective) for this kind of decision-making.

Nevertheless, all the evidence reported so far seems to approach value based choice decisions for social dilemmas as if social preferences were stable or in some cases, only outcome dependent. This view neglects the fact that (at least in real life) these decisions are always taken in the context of the potential decisions of other humans. That is, a second human agent (or player) might have an effect in the subject’s cognitive processes, thus making the analyses so far, somehow unrealistic. One cannot claim an individual’s choice is intuitive or reflective without considering the context of the social interaction under which the choice is made. A decision about whether to start running or not can be considered as intuitive or not depending on whether one is in front of a lion or meets a friend on the corridor, and so is the strength of preference and the expected outcome from the decision to run. Being more specific, in this kind of games, we believe that the social information gathered from the environment, in particular the information one can extract from the other human participant in the interaction, is an important variable guiding decision time and the factual decision to cooperate or not.

\(^1\) The way the authors measure this variable is by subtracting from the individual mean-median RT for cooperation the individual mean-median RT for non-cooperation. This operation will give the “general” preference towards one behavior or another.
The literature regarding the neural systems that mediate human perceptual decision-making has shown that when making a binary decision, subjects accumulate sensory information until they reach the response criterion. Once this boundary is reached, the decision process ends and a response is elicited, meaning sensory information is interpreted and translated into behavior (Heekeren et al., 2008). In a social interaction the relevant information would consist of the potential decision choices and the potential outcomes given a particular environment. When choosing between one action or another, this information is crucially going to drive the time it takes to reach a decision, and the decision itself.

Because by definition outcomes in a social interaction are dependent on both players’ decisions, we hypothesize the most relevant information for the subject is what s/he thinks/knows the other person will do. Therefore, individual efforts to extract and process the available information regarding the others’ potential actions will be made. Even in the case where there is no immediate information available, subjects involved in the interaction will make use of previous experiences, beliefs and heuristics about the social environment in which they are making a decision to reach a response criterion (Kahneman et al, 2002; Tomasello et al 2005).

So far studies using RTs for social dilemmas have not considered the important source of variation in preferences and behavior given by this perception/belief of the social environment, including an estimate of the intentions of other players. Tomasello et al 2005 labeled this as perceived intentionality. According to Tomasello et al, when acting towards a goal in a social interaction the actor perceives and understands the reference value or goal of the system in which s/he takes action; s/he understands that both actors have the ability to change the environment; and s/he has the ability to perceive when the state of the environment matches her reference goal. The subject will then act as to bring to the real environment his pursued goal, understanding that the other will perceive and evaluate reality accommodating his action plan also in pursuit of his goal. This has not been considered so far in the analysis of decision-making in social-dilemmas as a potential source of variability affecting reaction times.

In this particular study we want to fill this gap in the literature by exploring how this perceived intention of the other influences decision times (RTs) and individual choice in a cooperation dilemma. To address this question we use a discrete Prisoner’s Dilemma for two players where we induce a conflict between self-interest and social welfare. In order to manipulate the perceived intention of the other player, the subject was informed of the likely strategy adopted by the second player (probability of cooperation), prior to being presented with the decision screen displaying the two possible choices. We recorded the time it took the subject to reach a decision between the two options, one being more cooperative and one less cooperative.

We hypothesize that the cognitive decision process to reach a decision and in particular RTs will be strongly driven by social information about others (perceived) intentions, diminishing the effects of the monetary incentives previously found in past studies (Fehr et al. 2015). In this scenario, we claim RTs would be a good measure for intuitive and reflective behaviors in social dilemmas only that specific perceived intentions of the other
will induce different cognitive schemas to guide decision-making. Whereas some social priors will lead intuitive (hence fast) cooperation reactions, some others will instead induce cooperation to be the result of reflective (hence slow) cognitive processes.

Specifically, we predict that in this cooperation dilemma for a two-individual interaction the perception of a selfish (non cooperative) intentionality of the other will incentivize a cognitive schema where choosing the less cooperative option is intuitive whereas we will expect that a positive (cooperative) intentionality of the other will lead to a cognitive schema where cooperative decisions are more intuitive. This modulation of social context can, of course, sit on top of generic preferences toward cooperating or non-cooperating, overall.

For the decision to cooperate we anticipate a negative linear relationship between RTs and the intentionality of the other, meaning the more cooperative the strategy of the other, the less time it will take the individual to reach a cooperative decision.

Inversely, for the decision defect (not to cooperate), we anticipate a positive linear relationship between RTs and intentionality of the other, meaning the more cooperative the strategy of the other is perceived to be, the longer it takes for the individual to reach a non-cooperative decision.

2. Experimental Design

2.1 The experiment

In a two players Prisoner’s Dilemma, subjects (n=50) were shown for 105 trials two different amounts of money and they had to choose one. If both players (subject and a simulated player) chose the same amount of money, both got that amount, otherwise the player who chose the smaller amount got more money (Lprice+P) and the one who chose the larger amount got less money (Lprice – P) (Figure 1).

This dilemma was framed in the following way so to make it more realistic and easier to understand for the participants and:

“Two travelers come home from a remote island where they bought identical souvenirs. At their arrival the passengers find out their souvenirs have been destroyed. The airline will have to pay a compensation equal to the price of the souvenir to the passengers. For the airline it is impossible to know the exact price of the souvenir, nonetheless the airline has estimated two possible prices from a range in which the price may fall.”
On the search for the honest price of the souvenir the airline separates the 2 travelers so they cannot confabulate, and asks them to choose separately one of the two prices for the souvenir.

If both report the same price it would be reasonable to assume that both are saying the truth and the airline will reimburse this quantity of money to each of the passengers.

If the passengers report different prices, the airline will take the lowest price as the true cost of the souvenir and each passenger will receive this amount as compensation. Nevertheless, a discount on the amount of the compensation is going to be made to the passenger that reported the Highest Price as a Penalty. This amount will be given to the passenger that reported the Lowest Price as a Reward for his “honesty”.

Given this framework we will interpret the decisions in our analysis as follows:

- **Choosing the High Price will be taken as a decision to Cooperate**: This decision of the game for both players and therefore reflects that the participant cares about the second participant and is acting upon a egalitarian distribution of the gains of the game.
- **Choosing the Low Price will be taken as a decision to Defect**: This decision evidence an intention to maximize the self-compensation by leaving the other participant in a worst situation.

### 2.2 Conditions of interest (Figure 2)

Every round, before the subjects had to make a decision between the two different amounts of money, they were shown for two seconds the strategy adopted by a simulated second player as a ratio; the number of trials the other player was going to cooperate by choosing the Highest Price. By manipulating the strategy of the simulated player we attempted to manipulate the subjects perceived intentionality of cooperation from the other player. (From here onwards we will refer to this variable as intentionality). The prices (low and high) associated to each decision, cooperate or defect were also manipulated in order to vary the attractiveness of each decision in terms of monetary gains.

Here we present a description the manipulations used as controls in the design to create different tradeoffs intervening in the subject’s decision to cooperate:

- **Intentionalities:**
  5 different intentionalities (other player strategies) were shown randomly to the subject as if he was matched with different
players. The presented strategies were: “Never choosing the High Price” (never cooperating), “Always choosing the High Price” (always cooperating), and 20%, 50% and 80% of the rounds choosing the High Price (cooperating).

- Incentives:
The monetary outcomes (high and low prices) between rounds were manipulated so to generate three kind of situations: High-benefit’ rounds in which cooperation was costly and the subject had to give up a lot of money (= 6€) so himself and the other player could get the high price, ‘Low-benefit’ rounds in which if the subject decided to cooperate he had to give up very little money (= 1€) so both players could get the high price, and rounds in between this two in which the cooperation decision implied giving up 3€. (See the model used to build payoff structure in the next section)

2.3 Model and Payoff Structure

The expected outcomes for the subjects choosing the High Price of the souvenir (cooperating) and choosing the Low Price (defecting) would be respectively:

\[
U_{\text{High Price}} = H_{\text{price}} \cdot X_A + (L_{\text{price}} - P) (1 - X_A)
\]

\[
U_{\text{Low Price}} = (L_{\text{price}} + P) \cdot X_A + (L_{\text{price}}) (1 - X_A)
\]

where \( H_{\text{price}} \) refers to choosing the High Price (cooperate), \( L_{\text{price}} \) refers to choosing the Low Price (defect), \( X_A \) is the probability of Player A choosing the High Price (probability of Player A cooperating), and \( P \) is the “penalty” imposed if they chose different prices.

Player B choosing to Cooperate (choosing the High Price) would be a dominant strategy if for all sufficiently small values of \( X_A > 0 \), the expected Utility of choosing the High Price would be bigger than the expected Utility of choosing the Low Price:

\( H_{\text{price}} > L_{\text{price}} + P \)

This will never happen in this game because within this condition there is no cooperation dilemma. Thus, the dominant strategy of the game in terms of monetary gains is to always defect.

The payment structure of the game was designed considering three different scenarios where \( (L_{\text{price}} + P) - H_{\text{price}} \neq 0 \);

The bigger the difference between \( (L_{\text{price}} + P) \) and \( H_{\text{price}} \), the less incentives there are for an individual to cooperate. We varied this incentives presenting ‘high-benefit’, ‘median-benefit’ and ‘low-benefit’ trials \((L_{\text{price}} + P - H_{\text{price}}) = 6 \; 3 \; 1 \) euros respectively)
3. Materials and Methodology (Figure 3)

The experiment was programmed in Psychotoolbox (Matlab R2015B) and RTs were recorded in the decision screen.
- Subjects came to the laboratory in groups and each was assigned a cabin with a computer to play.
- Once in the cabin, participants were assigned a type of player: A or B (all participants were assigned B since A - the player that had to show his strategy - was simulated by the computer) (See Supplementary material 11.1 for the instructions)
- The game was presented as a Coordination Game avoiding strong language frames interfering in the decision subjects made.
- Instructions were then presented to the subjects and control questions were done at the end to ensure they had understood the game.
- 105 different trials/decisions were presented to the 50 subjects.
- Each trial, before presenting the decision screen, the strategy (intentionality) of the second player (player A) was shown to the subject as a ratio in a colored rectangle. Color in the rectangle showed the number of trials the second player had chosen the High allocation of money (to cooperate).
- Subjects chose the preferred amount of money with the numbers 1 and 3 from the keyboard.
- Reaction times were measured in the Decision Screen and no time restriction over the decision was imposed.
- Concluded the game, one trial was randomly selected at the end of the game to be paid. The final payment was dependent on the decision made by both, Player A (simulated) and the subject.
- Participants could earn between 0 to 20 euros in the game and trials involving losses where omitted to avoid loss aversion effects.

4. The Data

In this section we will describe the collected data.

The number of subjects used in the analysis was N=49 (one subject was excluded because RTs were not correctly recorded). Each participant played during 105 rounds but the two first rounds corresponded to practice rounds. Thus we have 103 decisions and the corresponding RTs for each decision.
Variables used in the analysis:

Independent Variables:
- **Decision**: Subjects' decision to Cooperate (Choosing High price) or Defect (Choosing Low Price) each round
- **Decision Time (RTs)**: the time it took the subjects to reach his/her decision each round

Dependent Variables (controls)
- **Intentionality**: other player's strategy shown to the subjects each round, which gave way to 5 different conditions: 0%, 20%, 50%, 80% 100% of intention of cooperation (choosing the High Price).
- **Incentives**: difference in monetary gains between choosing to cooperate or defect. 3 different conditions: High = 6 euros, Medium = 3 euros, Low = 1 euro.

Summary statistics are shown in Table 1 describing the behavioral results (RTs and Decisions) for the different intentionalities and incentives. RTs were measured in seconds.

Table 1. Summary Statistics

<table>
<thead>
<tr>
<th>Intentionality</th>
<th>Mean</th>
<th>Std. Err.</th>
<th>Cooperate</th>
<th>Defect</th>
<th>Observ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.242</td>
<td>0.062</td>
<td>43</td>
<td>989</td>
<td>1029</td>
</tr>
<tr>
<td>20</td>
<td>3.223</td>
<td>0.117</td>
<td>234</td>
<td>791</td>
<td>1029</td>
</tr>
<tr>
<td>50</td>
<td>2.558</td>
<td>0.083</td>
<td>84</td>
<td>949</td>
<td>1029</td>
</tr>
<tr>
<td>80</td>
<td>3.281</td>
<td>0.120</td>
<td>256</td>
<td>731</td>
<td>987</td>
</tr>
<tr>
<td>100</td>
<td>3.127</td>
<td>0.144</td>
<td>437</td>
<td>537</td>
<td>973</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Incentive</th>
<th>Mean</th>
<th>Std. Err.</th>
<th>Cooperate</th>
<th>Defect</th>
<th>Observ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.115</td>
<td>0.092</td>
<td>388</td>
<td>1273</td>
<td>1666</td>
</tr>
<tr>
<td>3</td>
<td>2.726</td>
<td>0.068</td>
<td>347</td>
<td>1374</td>
<td>1715</td>
</tr>
<tr>
<td>6</td>
<td>2.804</td>
<td>0.092</td>
<td>319</td>
<td>1350</td>
<td>1666</td>
</tr>
</tbody>
</table>

5. Proposed Analysis

We will present a statistical analysis of the data from the experiment in order to evaluate the hypotheses stated in the introduction. Thus, it points to the following two hypotheses:

I. *Perceived intentionality of the other player* drives response times (RTs), but it should present a different relationship with reaction times for cooperative versus not cooperative decisions. In particular, as the intentionality of the other subject cooperating is higher, the decision to cooperate will take less time and the decision to defect more time, whereas the reverse will be true for the defect decisions.

II. While Fehr et al (2015) have argued that incentives and strength of preference are the main factor driving reaction times, previous literature has omitted the perceived intentionality of the other player as a key variable in the decision making process and the time it takes a subject to reach a decision. The analysis here presented aims
to fill this gap by showing that strength of preference and incentives play a small role once perceived intentionality of the other is taken into account in the analysis of RTs.

In order to prove these two hypotheses, the results section will be divided into three sections:

(i) We will show that both, incentives and the intentionality of cooperation affect the decision to cooperate in the expected direction in a cooperation dilemma. These results support the design of the experimental setup and motivate to further explore reaction times.

(ii) We will study how RTs respond to the perceived intentionality of the other player and show that while incentives have an effect over RTs, their effect is marginal once the variable intentionality is included. Even though our analysis can be underpowered for certain values of probability (See power analysis in Appendix 10.1), the general conclusion is that the expected patterns emerge. Interestingly, when subjects face a 50% intention of cooperation the main pattern is reversed showing a role of uncertainty that should be explored in future research.

(iii) Fehr et al 2015 results assert that after controlling for strength of preference (extent in which individuals have a general preference towards cooperation or selfishness) they find no difference in RTs between the two decisions for each level of incentive. This argument supports their conclusion of variable incentives driving RTs, and is further used to claim RTs are not a good measure to infer cognitive processes corresponding to intuition or reflection for cooperation decisions. In this last section we will control for individual preferences by including in our analysis individual fixed effects. Controlling for this fixed preferences we will further examine the effects of our proposed variable intentionality in RTs for each decision finding significant differences in RTs for cooperation and defection. Moreover we will show evidence corroborating cooperation and defection behaviors are not only different within intentionalities but have opposite RTs tendencies.

6. Results

i. The first set of results corresponds to the determinants of the decision to cooperate (i.e., note this refers to the probability to cooperate, not the RTs). For simplicity, in the analysis we are going to refer to the variable perceived intentionality of the other player as intentionality (to cooperate).

In the first part of the analysis, we estimate the following equation for the probability of a subject to cooperate:

\[ \text{cooperate} = f(\beta_0 + \sum_{j=1,3} y_j \text{incentives}_j + \sum_{d=\{0,20,80,100\}} \rho_d \text{intentionality} ) \]

Where cooperate corresponds to a dummy variable of whether the subject decides or not to
cooperate, incentives to a set of incentive dummies on the experimental set-up, and intentionality to the set of information given of the other player’s behavior described before. We estimate this model using a Probit regression that accounts for the probability of a discrete change of the subject deciding whether to cooperate or not given the different monetary incentives and intentionals. Coefficients from the regression are plotted in Graphs 1 & 2 (find table from the estimated regression in Appendix section 10.2).

Graph 1. Intentionality

Graph 2. Incentives

In Graph 1 the coefficients from the regression are plotted showing the changes in probability of a subject deciding to cooperate for each level of intentionality. Correspondingly, coefficients for each level of incentive were plotted in Graph 2. The middle variables for both intentionality and incentive were dropped from the model because of collinearity (50% of intention of the other cooperating and incentive=3 correspondingly) so the analysis for each variable is done against dropped one. Thus, the x-axis in both graphs corresponds then to the dropped variables.

In general, we can interpret the condition of the 50% of intention as having no information since the other shown intentions suggest or signal cooperative or non-cooperative strategies followed by the other player. Graph 1 shows the changes in probability of the subject’s cooperating from the no information condition (50% of intention) to the conditions were a clear intention is signaled (0%, 20%, 80%, and 100%). The x-axis corresponds to the 50% of intention and the y-axis displays the change in probability. Results displayed in Graph 1 show that compared with the condition 50% of intention, more cooperative intentions (80% and 100%) increase (in 23 and 41.8 percentage points respectively) the subjects’ likelihood of cooperating. As predicted a 0% of intention (never cooperating) decreases (in 7.6 percentage points) the probability of the subjects cooperating in comparison to being shown a 50% of intention. Note though that for the 20% of intention we see an increase (of 19.6 percentage points) in the probability of the subjects cooperating even though the 20% condition is informing the subject of a less cooperative strategy from the other player. This suggests subjects in the no information condition are less prone to cooperate that in cases where the other player signals consistent, albeit relatively non-cooperative behaviors like the 20% of intention to cooperate. This result would account for certain irrationality in
behavior since 50% of intention corresponds to a more cooperative strategy. We will further examine this result in the discussion section.

Graph 2 displays the coefficients for a change in the probability of a subject’s decision to cooperate for each level of incentives compared to having a medium incentive (= 3 euros). First note that the changes in the probability of a subject’s decision to cooperate are considerably small if compared with the coefficients for the effects of the different intentionalities, suggesting a larger effect of the variable intentionality. These coefficients inform us that compared to having the medium monetary incentive (=3 euros), having the smallest monetary incentive (=1 euro) increases (in 4 percentage points) the probability of the subjects deciding to cooperate. The coefficient for the high incentive (=6 euros) is very small and not statistically different from the medium incentive (=3 euros), this implying subjects act similarly when the incentive is medium or high, and are more prone to cooperate when the incentive is low.

Results from this regression show the expected direction of coefficients for the variable intentionality and incentive; the higher the intention from the other to cooperate the more probable it is that the subject decides to cooperate and the smaller the incentive (difference in utility between cooperating or not) the more probable it is that the subject cooperates (hence, replicating the results of Fehr et al 2015). Note though, that coefficients for the intention variables are considerably higher suggesting a larger effect of intention variables than incentive ones on decision to whether cooperate or not. These results support the consistency of the experimental design and we can move forward to explore how these two variables affect decision time, and to further study the 50% of intention condition that appears as a case of biased behavior.

ii. In this second set of analysis we explore how reaction times are affected by the intentionality of the other cooperating for the decision to cooperate or defect in order to evaluate the hypotheses presented in the introduction.

For this matter, we will start by looking at the probability density function of RTs for the decision to cooperate and to defect separately since outcomes associated to each decision are different. We predict that as the variable intentionality of the other to cooperate augments, RTs associated to each decision will follow opposite tendencies and this should be evidenced in their distribution on time. The probability distribution for each decision will describe the relative likelihood from RTs to take certain values given a particular intentionality. For this matter RTs were log10 transformed as done in reaction times literature to account for a right-skewed distribution that helps in visualizing their distribution better.

Figure 4 shows, first, a graph with the overall probability distribution of RTs for the decision to cooperate (blue) and the decision to defect (red). Following this first graph, distributions of RTs for each level of intention are shown. The x-axis corresponds to the log10 transformed values for RTs and the y-axis to its density (the proportion of responses
concentrated in that range of time).

**Figure 4.** Probability distributions of RTs

The first graph displays a normal density distribution for both decisions - to cooperate and to defect across all conditions. Next, the density distribution is shown separately by level of *intentionality*. The plotted RTs distributions for each response category evidence that from the 0% condition to the 100% condition, distributions for both behaviors (cooperation and defection) invert their pattern in terms of RTs. When the *intentionality* of cooperation is low (0%, 20% and even 50% conditions) defection decisions are concentrated in faster RTs than cooperative ones. Nonetheless as the *intentionality* of cooperation augments the density distribution for defection starts shifting right towards slower RTs. Contrariwise, as the *intentionality* of cooperation augments the RTs density distribution for cooperative behaviors shifts left concentrating in slower RTs and ending up being faster than defection decisions. (Note though that for some decisions the sample is very small since the data presents overall more defecting decisions n= 3895 than cooperative ones n=1152).

As to further explore the effect of the proposed variable *intentionality* in RTs, we now present the means for each decision. The decision to cooperate is presented in blue and the decision to defect in red with the corresponding confidence intervals. Both decisions, cooperate and defect, are presented in the same graph and sorted by *intentions* (graph 3) and by *incentives* (graph 4) to see their effect in mean RTs and if the effects follow any trend. The x-axis shows the level of *intention to cooperate* in Graph 3 and the level of *incentive* in Graph 4. The y-axis in both graphs presents RTs.
Decisions to defect are overall faster (mean= 2.79s) than decisions to cooperate (mean= 3.19s). Nonetheless in Graph 3 where mean RTs for cooperation and defection are plotted by intentionality we can see that RTs associated to both decisions are strongly dependent on this variable and even for some intentionalities cooperation decisions can be faster than defection ones.

Coherent with the results from the density distributions by intentionality, the data for the means in Graph 3 suggest that decisions to cooperate and defect have opposite trends as the intentionality to cooperate augments. For the decision to cooperate the relationship between intentionality and RTs is negative; as the strategy of the other player becomes more cooperative, it takes less time to the subjects to reach the decision to cooperate. Inversely, for the decision to defect we have a positive relationship between intentionality and RTs; as the other player becomes more cooperative, it takes more time for the subjects to defect. The case of 50% of intention seems to be a special case, as the RTs results do not seem to abide to the described RTs trends for both decision categories as a function of the other players intention and will be further discussed with more evidence.

Graph 4 displaying the mean RTs by incentives evidence very similar mean RTs within and between levels of incentive with vast standard deviations for each level. No conclusion can be therefore drawn for this variable, and so far differences in means between levels of intentionalities suggest this is the variable mainly driving RTs.

In order to quantify the strength of the relationship between RTs and intentionality we estimate two models looking at RTs as a function of choice type (to cooperate or defect). The first model accounts for the effect of intentionality in RTs alone, and the second one accounts for the effect of the interaction of both variables, incentives and intentionality.

The following models were estimated with a multiple linear regression for both decisions:

\[ \text{Log}_Rts = (\Sigma_{d=0,20,50,80,100}) \rho_d \text{ intentionality} ; \]
if $\text{Cooperate} = 0 \text{ and } 1$

\[ (3) \quad \text{Log}_\text{Rts} = \left( \sum_{d=\{0,20,50,80,100\}} \rho_{d*i} \text{intentionality} * \text{incentive}_j \right) \]

if $\text{Cooperate} = 0 \text{ and } 1$

Where $\text{Log}_\text{Rts}$ corresponds to the RTs log10 transformed in both models, $\text{intentionality}_d$ to the set of intention dummies for each condition in model (2) and $\text{intentionality}^*\text{incentives}$ refers to the interaction between each dummy for $\text{intentionality}$ and for $\text{incentives}$ in model (3). This last model will give the marginal effect for each interaction $\text{intentionality}^*\text{incentive}$.

Coefficients obtained from the regression capturing the effect of each variable in the RTs are plotted in Graphs 5a and 5b (model 2) and Graphs 6a and 6b (model 3).

The estimation of model (2) regarding the effect of the variable $\text{intentionality}$ in RTs both decisions display statistically significant coefficients for each level of intention (Appendix 10.3). Coefficients and confidence intervals are plotted in Graphs 5a and 5b. In both graphs the x-axis corresponds to the level of intention to cooperate from the other player and the y-axis to the coefficients of the regression capturing the effect of $\text{intentionality}$ in RTs.

**Graph 5a.** Decision to Defect

**Graph 5b.** Decision to Cooperate

Results from this first regression evidence that the effect of the variable $\text{intentionality}$ on RTs is not linear. However, if we disregard the 50% condition also referred previously as the no information condition from the analysis, results from the regression confirm a negative relationship between $\text{intentionality}$ and RTs for the decision to cooperate. It must be said though that given that the number of decisions to cooperate in the sample are considerably small, there is substantial variability in data and not enough power to assert something about this decision. On the contrary, for the decision to defect we have enough power to assert that there is a positive relationship between $\text{intentionality}$ and RTs; as the other player becomes more cooperative, it takes more time for subjects to defect.
The 50% condition might be nonetheless the most interesting case. Although there appears to be a trend for each decision as the intention to cooperate increases, for both decisions to cooperate and defect, the 50% of intention condition seems to invert the tendency RTs follow. Subjects in the 50% condition are not only less cooperative compared to the 20% condition as seen in results section (i), but are also faster to defect and slower to cooperate if compared with the 20% condition reverting the RTs tendencies for both decisions. This result confirms a bias in decision that will be further considered in the discussion section.

Results from the second regression (model 3) regarding the interaction between intentionality and incentives are plotted in Graphs 6a and 6b (find table for the regression results in Appendix 10.4). Here the x-axis represents each interaction between intentionality and incentives used in the experimental design, and the y-axis the obtained coefficients for the effects of each interaction in RTs. Coefficients for all interactions are statically significant.

As it can be seen in the Graphs 6a and 6b, the previous tendency shown for RTs in relation to intentionality are maintained, and are opposite for the decision to cooperate and defect. These results confirm the variable intentionality drive RTs since larger differences of RTs are found across rather than within intentionalities suggesting a small role of incentives with no big differences between them within each intentionality. For some decisions to defect we find the reported tendency of RTs being faster when incentives are bigger found by Fehr and colleges but this tendency is not maintained across all intentionalities. Contra intuitively this tendency is also found for the decision to cooperate suggesting the bigger the utility difference the faster the person would be to cooperate. Nevertheless as stated before we lack power in the data to state conclusions about this decision, and our interest variable intentionality behaves as predicted. From these results we can claim that the proposed variable intentionality is predominantly the variable driving RTs after controlling by the level of incentives that appear here to have a marginal role.
So far we have demonstrated that intentionality has an important effect over RTs, and that this effects differ for cooperative and non-cooperative decisions as it was hypothesized in the introduction and validated with the previous set of results. Moreover, our previous analysis has demonstrated incentives have a marginal role as an explicative variable driving RTs once perceived intentionality is taken into account in the analysis. In this last set of results we want to explore the role of intentionality in driving RTs including in our analysis a proxy for the variable strength of preference proposed by Fehr et al 2015. Reported results by Fehr and colleagues find no difference in RTs between selfish and pro-social decisions for each level of incentive (high, medium and low) when controlling by the strength of preference. In this last section we will show we do find differences in RTs between the decision to cooperate and defect within the same level of intentionalities after controlling for strength of preference. More over show here robust evidence of this two behaviors having an opposite tendency in RTs driven by changes in intentionality.

Fehr and colleagues used the difference in individual mean-median RTs for cooperative and selfish decisions as a proxy of the individual preference towards one behavior or another. In this experimental design we manipulate intentions and incentives at the same time making it impossible to disentangle preference for a particular decision since, as we have seen so far, preferences for one behavior or another will be mediated by the perceived intentionality. Nevertheless we believe is reasonable that, independent of the intentionality of the other, some people might be more prone to always cooperate/defect and so far we are not capturing this in our analysis. Thus we will account for the strength of preferences using individual fixed effects, since this technique allow us to control for all individual characteristics that remain constant across the experiment by assuming that the preference for cooperation or defection is a variable at the individual level.

The estimation we do is the following:

\[
\log_{10}\text{RTs} = (\beta_0 + \sum_{d=0,20,80,100} \rho_d \text{intentionality} + \varnothing \text{Fe}) \times \begin{cases} 1 & \text{if Cooperate}=0 \text{ and } =1 \\ 0 & \text{otherwise} \end{cases}
\]

Where \(\log_{10}\text{RTs}\) corresponds to the RTs log10 transformed, and \(\text{intentionality}_d\) to the set of intention dummies for each condition, and \(\text{Fe}\) all individual characteristics that remain constant across the experiment. The model is estimated for both, the decision to cooperate and defect and results are plotted in Graph 7. The x-axis corresponds to the level of intention and the y-axis for the coefficients of the effects of intentionality in RTs after controlling by individual fixed effects. Cooperative decisions correspond to the blue dots and defection decisions to the red dots.
The results obtained are very similar to what we had seen so far. RTs for cooperation and defection follow opposite tendencies conform intentionality changes, and it is especially evident for the 0%, 50% and 100% conditions. The individual fixed effects do not change these results and it can be concluded that RTs are different for cooperation and defection behaviors but none could be claimed to be intuitive and fast or reflexive and slow. As we had predicted different cognitive schemas seem to be incentivized depending on the perceived intentionality of the other and thus choosing to cooperate/defect can be either fast or slow independent of existing fixed characteristics from the subjects that could be thought to affect decision times.

Finally in order be able to state RTs for both decisions follow a different and opposite tendency driven by the variable intentionality as we have above suggested, a linear regression for RTs dependent on intentionality, incentives, the decision to cooperate, and the interaction between this last two variables (cooperation and intentionality) must be conducted. We again include a control for individual fixed effects.

The model we estimate is the following:

\[ \text{Log}_\text{Rts} \left( \beta_0 + \beta_1 \text{intentionality}_d + \beta_2 \text{incentive} + \beta_3 \text{cooperate} + \beta_4 \text{intention} \times \text{cooperate} + \phi \text{Fe} \right) \]

Where Log_Rts, intentionality and incentives have been already defined, cooperate corresponds to a dummy for the decision to cooperate, and the interaction intention*cooperate will describe the opposite tendency in RTs for cooperation and defection if it exists and Fe refers to the individual fixed effects that captures the strength of preference of each subject. Results for the estimation are shown in Table 2.
The results obtained from the regression show the expected direction for the three first variables: a positive relationship between intention and RTs that had been previously reported, the larger the intention to cooperate the more time it takes to reach a decision. For the variable incentive a negative coefficient suggesting the bigger the incentive the faster the subjects to reach a decision, and for cooperation (the decision to cooperate) a positive coefficient suggesting more cooperative intentions trigger longer RTs as it does the decision to cooperate in comparison with the decision to defect.

Our interest variable in this estimation corresponds to the interaction of the variables intention and cooperation. The coefficient we get is negative and is notifying that for the cooperation decisions, as the intentionality of cooperation increases, RTs decrease. This result contradicts the effects reported by coefficients for intentionality and cooperation alone (increases in intentionality alone triggers larger RTs, and so is the decision to cooperate in comparison with the decision to defect). Thus, what a negative coefficient for the interaction allows us to conclude is that there exists an opposite trend in RTs triggered by intentionality that makes RTs for cooperative behaviors decrease as the intentionality of cooperation of the other player increases opposite to what happens with defection behaviors.

7. Discussion

Previous literature in social decision-making for cooperation dilemmas has dismissed RTs as a good measure to infer cognitive processes arguing this variable is driven rather by fixed preferences, cognitive conflict (corner responses being faster for a continuum choice set) or similarity in choice (the more similar two options are in terms of outcome the longer it takes to reach a decision). This literature has mainly focused in experiments where...
monetary rewards and time available for subjects to reach a decision were manipulated while they were considered the main variables driving RTs and choice.

We consider this previous emphasis is biased since in social dilemmas outcomes are by definition interdependent. Thus an accurate and more realistic analysis of social behavior would require taking into account the potential role of a second participant in the individual’s cognitive processes. The idea behind this argument is that in order to reach any decision, subjects extract and accumulate relevant sensory information from the environment that can help an appropriate choice until they reach a response. In a social interaction, aside from the information regarding the potential gains, we hypothesized subjects would assess the other participants’ potential actions as relevant information in order to reach a decision. Furthermore, we inferred this information gathered from other human participant would prompt different cognitive processes that could be reflected in decision times.

In this study, we used a cooperation dilemma were the subject’s perception of the second human player’s intention to cooperate in the game was manipulated. With the results obtained we were able to show that the subjects’ perceived intentionality of cooperation of the other player was the most relevant variable driving decision times for the decision to whether cooperate or defect. Moreover, our results show robust evidence that when presented a binary choice, RTs for both decisions, to cooperate and to defect, are not only different but follow opposite trends given the same intentionalities: cooperative behaviors are faster as the perceived intention of the other to cooperate is larger and opposed to this, defection behaviors take more time as the presented intention to cooperate is larger.

Remarkably, the effects found in this study for the monetary incentives in decision times are marginal. This variable was manipulated so to make the choices to cooperate and defect very similar in terms of monetary gains, or very different. Although results obtained replicate the previous evidence on mean RTs tending to increase as choices become more similar for the decision to defect, results also evidence that the perceived intention of cooperation has overall bigger effects on RTs, sometimes even erasing the expected effect of incentives.

This result is particularly relevant since RTs being driven primarily by similarity of choice was asserted true for both perceptual and value-based decision making (Fehr et al 2015). While for perceptual and some decision-making tasks that do not involve social interactions as inter-temporal choice this may hold true, we have proven that in a social interaction, in particular in a cooperation dilemma, RTs are mainly driven by the perception of intention of the second participant different to what had been previously stated.

Furthermore, Fehr and colleagues incorporated in their analysis of RTs a variable that captured individual preference because they assumed that people have beforehand fixed individual preferences towards cooperation or selfishness. Their results show that after controlling for strength of preference between choice options there is no difference in RTs between selfish and cooperative decisions given a fixed level of incentives. In our analysis we control for these preferences using individual fixed effects that capture all individual
characteristics that remain constant across the experiments and we do find differences in RTs for cooperation and defection decisions between the same intentionalities. Furthermore, we find an inverse tendency in RTs for both behaviors, these results once again suggesting decisions to cooperate and defect are driven in an opposite fashion not by incentives, as previously proposed in the literature, but by intentionalities.

As to further explore the variables driving RTs we conducted a survey at the end of the study regarding the strategies used by the participants in the decision process. One of the questions explored what the subjects considered was the most relevant information to reach a decision (intentionality, incentives or both). Results confirm intentionality (the other participants strategy) to be the most relevant information for the subjects with a 37% of responses, 26% for both intentionality and incentives, and 21% for only incentives (the resting 16% regarded other variables). The majority of participants when asked to discriminate between relevant information in the decision process assessed the visual stimuli for other player’s intentionality of cooperating as the most relevant information to reach a decision in the cooperation dilemma. The information about incentives is regarded, but secondary.

Based on this evidence we confirm our hypothesis of intentionality mainly driving RTs and add evidence to the debate on whether cooperation/defection can be considered intuitive or reflective behaviors. Our contribution consists in demonstrating that for each level of intentionality, each behavior has an intuitive and fast response and a reflexive and slow one. Cooperative behaviors can be seen as intuitive and fast when cooperative behaviors from a second participant are perceived, whereas defection is intuitive and fast when non-cooperative behaviors are perceived and viceversa.

**Uncertainty and Irrational Behaviors:**
**What happens with the 50% intention of cooperation?**

So far we have discussed the general results for the effects of intentionality in RTs, which show opposite tendencies for the decisions to cooperate and defect, as predicted. Interestingly, in both cases, the condition of the 50% of intention to cooperate reverts the main RTs tendency that each behavior follows conform intention of cooperation increases: RTs for cooperative behaviors tend to decrease (and RTs for defection tend to increase) as the intention of cooperation increases but not for the 50% condition.

Overall, subjects in the 50% condition cooperated less than when shown a 20% intention of cooperation. From 1029 decisions recorded for the 50% intention of cooperation condition, 90.2 percent were to defect whereas for the less cooperative condition of 20% of intention of cooperation, decisions to defect represent the 75 percent of the decisions, this last condition being less cooperative than the 50% condition. This apparent inconsistency in choice holds not for both decisions and associated RTs. Defection decisions for the 50% condition are almost as fast as they are for the 0% condition (2.55 and 2.24 seconds correspondingly), whereas in the 20% condition the cognitive process to reach the decision
to defect takes in average longer (3.22 seconds). It would then appear that when faced with the 50% intention of cooperation, subjects default to a similar strategy as when faced with the 0% cooperation intention, where defection decisions are done fast and are by far the preponderant choice.

A possible explanation for this “biased” behavior in the 50% condition could be supported by a bias in perception. In the survey a question regarding how cooperative the subjects found the different strategy was included. The question explored how cooperative subjects perceived the different strategies from the other player (intentionalities) in a scale of 0 to 7, 0 being “not cooperative at all” and 7 “very cooperative”. The question was asked showing to the participants different strategies as the visual stimulus (colored rectangles) they had seen in the experiment. Subjects on average perceived the exact same level of cooperation from the other player in the 0% and the 50% condition (average rate equal to 3 in the proposed scale), whereas the 100% condition was graded on average with a 5.9. Moreover, 31.4 percent of the participants rated the 0% strategy as “very cooperative” (= 7), whereas this perception was reported only by 11.4 percent of participants for the 50% condition (and 74,3 percent of participants for the 100 % condition). This results advocate for a bias in perception; for some reason the 0% condition is rated by almost a third part of the subjects from the sample as “very cooperative”, and the average perception for the 0% and 50% condition is the same.

In order to interpret this “bias” in perception some considerations must be made. The 0% and the 100% conditions represented certainty situations where participants could directly calculate how much they would earn from taking one decision or another, thus certainty could be taken as a cooperative behavior. The other conditions (20%, 50% and 80%) represented the opposite; these conditions implied uncertainty and therefore calculations done could only account for expected outcomes given the presented intentionality. In particular, the 50% condition could be perceived as the “perfect uncertainty” or no information condition since all other strategies adopted by the second player signaled a cooperative or non-cooperative behavior. Nonetheless, with the data available, apart from reporting a bias in perception we unable to give further explanations for the underestimation of the 50% of intentionality of cooperation and the uncooperative and fast behavior it entails, more closely related to the 0% condition.

Alternative explanations for this bias in decision and the corresponding RTs can be found in the literature. The first explanation we will use will be taken from the biases in information processing and probability judgment developed by Tversky & Kahneman (1975, 1983) that demonstrate heuristic processes are made when interpreting probabilities. Although we did not explicitly used frames of probability of cooperation but ratios, decisions made in the 50% condition could have been biased towards the natural evaluation that the problem evokes, in this case, towards no cooperation that would be the default answer to maximize the monetary gains. As the subjects in the 50% condition had no information signaling where the other participant’s decision could be guided towards, they might have failed in appreciating the difference in probability from the 20% to the 50% condition.
This explanation could be further complemented with Cacioppo’s el al (1997, 1988) work on adaptative value of negativity bias that demonstrate that in a neutral context larger brain event-related potentials are found for negative events than positive, suggesting a negativity bias emerges in the evaluative system when a stimuli contains equal units of positive and negative inputs (see also Rozin & Royzman 2001 for a complete discussion on negativity bias). This last explanation seems to adjust to the obtained behavioral results in our experiment if we interpret the 50% condition as neutral stimuli containing equal probability units from cooperative and uncooperative behavior from the other participant. Following Cacioppo and colleagues we could then explain the behavioral results for the 50% condition by arguing this neutral stimulus triggered a negativity bias that is later confirmed with the negative perceptions reported in the survey. This negativity bias had not been previously explored in social interactions and thus emerges as a relevant contribution to interpret human behavior under uncertainty.

RTs a good measure to infer cognitive processes in social dilemmas?

In social dilemmas, in particular in cooperation dilemmas where there is no paradigm of accuracy, RTs had been studied to see whether cooperative behaviors are intuitive (fast) or reflective (slow) using various experimental designs. Results obtained so far gave contradictory evidence on RTs, and thus the latest literature on RTs for cooperation dilemmas assert RTs to be a bad measure to infer cognitive processes in social dilemmas (Fehr et al 2015; Rand el al 2015)

In this study, the differences found in RTs for cooperation and defection for the same sensory stimulus (intentionality of the other player) are reflecting the duration of some set of mental processes that are different for each behavior given a level of intentionality. The analysis done within and between conditions for responses (cooperate/defect) and the corresponding RTs corroborate the hypothesis of perceived intentionality being a sensitive indicator affecting the underlying individual cognitive process for one decision or another: cognitive processes for cooperation and defection take different times depending on the individual perception of cooperation from the other player. Consequently, the presented evidence corroborates RTs being a good proxy for intuitive or reflective processes in social dilemmas if the correct variables are taken into the analysis.

8. Conclusion

The main purpose of this research was to test in a cooperation dilemma, the effects a second human player has in the subjects’ cognitive processes in decision-making by looking at RTs. In order to do this we used a experimental design were we induced to the subjects a perception of the second humans player willingness to cooperate (intentionality), at the same time varying the tradeoffs of cooperation by changing the monetary rewards (incentives) associated to cooperation and defection.

The first hypothesis we had was that both, incentives and perceived intentionality of cooperation of the second participant had an effect in the subject’s decision to cooperate or defect. Results show a big positive effect for the variable intentionality in the probability of
the subjects cooperating: the larger the perceived intention of cooperation of the second player, the more prone is the subject to cooperate. An effect from the variable incentives in the probability of cooperation was also found, but marginal.

The second hypothesis we examined regarded the effect of the different intentionalities (of cooperation) in RTs; we expected cooperation and defection behaviors to follow opposite tendencies conform the perceived intentionality of cooperation was larger. Results found support this hypothesis showing that as the intentionality of cooperation augments defection behaviors take more time whereas cooperation behaviors take less. These results suggest that there are fast and intuitive as well as reflective and slow responses for both the decisions to cooperate and defect depending on the intention subjects perceive from the second human player. Once again we controlled for incentives but the effects found for this variable in RTs are marginal.

Finally, we wanted to test the effects of our proposed variable intentionality on RTs assuming there are beforehand individual preferences for cooperation or defection that could explain the variations in RTs. After controlling for individual fixed effects we still find the expected effects in RTs from the perceived intentionality of cooperation of the second player, and an opposite trend in RTs for cooperation and defection decisions as intentionality of cooperation augments.

The 3 general conclusions that we can draw from this research are then: 1. That cooperation and defection behaviors can be either intuitive or reflective depending on the intentionality of cooperation subjects perceive from the second participant. This intentionality presents itself to be the most relevant information for decision-making in a social interaction. 2. There is a negativity bias in perception when subjects are faced with uncertainty (here understood as the 50% condition where equal units of positive and negative inputs where shown). Participants fail to evaluate uncertainty perceiving it more negatively than how they perceive pure non-cooperative strategies that involve certainty. Therefore, responses for the 50% condition are more closely related in RTs and factual decisions to responses in the condition where no intentions of cooperation from the other (0% condition) are shown. 3. We demonstrate social information regarding other participants is taken by subjects as relevant information in the decision making process and we add evidence on RTs being a good measure to infer cognitive processes for social dilemmas if the correct variables are included in the analysis.
9. Bibliography

• Cacciopo et al (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. Psychological bulletin
• Cappelen et al., (2014). Fairness is intuitive. Institute of Economics, University of Copenhagen.
• Civai et al. (2010). Are irrational reactions to unfairness truly emotionally-driven? Dissociated behavioral and emotional responses in the Ultimatum Game task. Cognition
• Grimm et al. (2011) Let me sleep on it: Delay reduces rejection rates in ultimatum games. Economic Letters
• Heekeren et al. (2008) The neural systems that mediate human perceptual decision-making
• Kahneman et al. (1986). Fairness and the assumptions of economics. Journal of business
• Krajbich & Rangel (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions.
• Proceedings of the National Academy of Sciences, USA of the United States of America, 108, 13852–13857
• Lohse et al. (2014). Giving is a question of time: Response times and contributions to a real world public good. Discussion Paper Series, University of Heidelberg, Department of Economics N566
• Singer et al. (2006). Empathic neural responses are modulated by the perceived fairness of others. Nature
• Tricomi et al. (2010). Neural evidence for inequality-averse social preferences. Nature
• Tomasello el al. (2005) Understanding and sharing intentions: The origins of cultural cognition. Behavioral and Brain Sciences. 28, 675-735
10. Appendix

10.1 Power Analysis

<table>
<thead>
<tr>
<th>Level of intention</th>
<th>No cooperate</th>
<th>Cooperate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obs.</td>
<td>Mean</td>
</tr>
<tr>
<td>0</td>
<td>944</td>
<td>0.559</td>
</tr>
<tr>
<td>20</td>
<td>754</td>
<td>0.812</td>
</tr>
<tr>
<td>50</td>
<td>903</td>
<td>0.645</td>
</tr>
<tr>
<td>80</td>
<td>714</td>
<td>0.887</td>
</tr>
<tr>
<td>100</td>
<td>521</td>
<td>0.956</td>
</tr>
</tbody>
</table>

Power Analysis 1: Power to state if the difference in the data gathered for the decision to cooperate and defect is statistically different from zero for each level of intention.

Power Analysis 2: Power to state if the difference in the data gathered for the decision to defect is statistically different from level of intention 50 (information vs no information).

Power Analysis 3: Power to state if the difference in the data gathered for the decision to cooperate is statistically different from level of intention 50 (information vs no information).

10.2 Probit Model for the probability of cooperation given the different levels of intentionalities and incentives.

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) Cooperar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intention 0%</td>
<td>-0.326***</td>
</tr>
<tr>
<td>(0.0898)</td>
<td></td>
</tr>
<tr>
<td>Intention 20%</td>
<td>0.657***</td>
</tr>
<tr>
<td>(0.0714)</td>
<td></td>
</tr>
<tr>
<td>Intention 80%</td>
<td>0.754***</td>
</tr>
<tr>
<td>(0.0712)</td>
<td></td>
</tr>
<tr>
<td>Intention 100%</td>
<td>1.278***</td>
</tr>
<tr>
<td>(0.0697)</td>
<td></td>
</tr>
<tr>
<td>Incentive 1</td>
<td>0.153***</td>
</tr>
<tr>
<td>(0.0518)</td>
<td></td>
</tr>
<tr>
<td>Incentive 6</td>
<td>-0.000841</td>
</tr>
<tr>
<td>(0.0527)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-1.451***</td>
</tr>
<tr>
<td>(0.0646)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>5,051</td>
</tr>
</tbody>
</table>
| Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

10.3 Linear Regression for Log RTs and the different levels of intentionalities (Decision to cooperate and defect).

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) Cooperate</th>
<th>(2) Defect</th>
</tr>
</thead>
<tbody>
<tr>
<td>log_rt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intention 0%</td>
<td>1.214***</td>
<td>0.555***</td>
</tr>
<tr>
<td>(0.127)</td>
<td>(0.0227)</td>
<td></td>
</tr>
<tr>
<td>Intention 20%</td>
<td>0.866***</td>
<td>0.800***</td>
</tr>
<tr>
<td>(0.0544)</td>
<td>(0.0254)</td>
<td></td>
</tr>
<tr>
<td>Intention 50%</td>
<td>0.990***</td>
<td>0.632***</td>
</tr>
<tr>
<td>(0.0905)</td>
<td>(0.0232)</td>
<td></td>
</tr>
<tr>
<td>Intention 80%</td>
<td>0.803***</td>
<td>0.855***</td>
</tr>
<tr>
<td>(0.0519)</td>
<td>(0.0264)</td>
<td></td>
</tr>
<tr>
<td>Intention 100%</td>
<td>0.629***</td>
<td>0.866***</td>
</tr>
<tr>
<td>(0.0397)</td>
<td>(0.0311)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>1,053</td>
<td>3,895</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.477</td>
<td>0.516</td>
</tr>
</tbody>
</table>

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
### 10.4 Linear Regression for Log_RTs and the interaction for *intentionalities*\(^*\) *incentives*

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) Cooperate</th>
<th>(2) Defect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inten0_Inc1</td>
<td>1.264*** (0.411)</td>
<td>0.516*** (0.0388)</td>
</tr>
<tr>
<td>Inten0_Inc3</td>
<td>1.476*** (0.212)</td>
<td>0.582*** (0.0392)</td>
</tr>
<tr>
<td>Inten0_Inc6</td>
<td>1.043*** (0.168)</td>
<td>0.568*** (0.0399)</td>
</tr>
<tr>
<td>Inten20_Inc1</td>
<td>1.058*** (0.0881)</td>
<td>0.820*** (0.0452)</td>
</tr>
<tr>
<td>Inten20_Inc3</td>
<td>0.806*** (0.0996)</td>
<td>0.764*** (0.0430)</td>
</tr>
<tr>
<td>Inten20_Inc6</td>
<td>0.706*** (0.0930)</td>
<td>0.820*** (0.0439)</td>
</tr>
<tr>
<td>Inten50_Inc1</td>
<td>1.111*** (0.128)</td>
<td>0.663*** (0.0410)</td>
</tr>
<tr>
<td>Inten50_Inc3</td>
<td>0.928*** (0.161)</td>
<td>0.622*** (0.0402)</td>
</tr>
<tr>
<td>Inten50_Inc6</td>
<td>0.792*** (0.199)</td>
<td>0.612*** (0.0393)</td>
</tr>
<tr>
<td>Inten80_Inc1</td>
<td>1.000*** (0.0847)</td>
<td>0.900*** (0.0455)</td>
</tr>
<tr>
<td>Inten80_Inc3</td>
<td>0.716*** (0.0907)</td>
<td>0.826*** (0.0460)</td>
</tr>
<tr>
<td>Inten80_Inc6</td>
<td>0.661*** (0.0918)</td>
<td>0.837*** (0.0460)</td>
</tr>
<tr>
<td>Inten100_Inc1</td>
<td>0.792*** (0.0645)</td>
<td>0.946*** (0.0629)</td>
</tr>
<tr>
<td>Inten100_Inc3</td>
<td>0.560*** (0.0658)</td>
<td>0.853*** (0.0498)</td>
</tr>
<tr>
<td>Inten100_Inc6</td>
<td>0.499*** (0.0753)</td>
<td>0.826*** (0.0513)</td>
</tr>
</tbody>
</table>

Observations: 1,053, 3,895

R-squared: 0.493, 0.517

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1
### 10.5 Fixed Effects Regression for RTs and *Intentionalities* (decisions to cooperate and defect)

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) Cooperate</th>
<th>(2) Defect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>log_rt</td>
</tr>
<tr>
<td>Intention 20%</td>
<td>-0.0903</td>
<td>0.261***</td>
</tr>
<tr>
<td></td>
<td>(0.104)</td>
<td>(0.0276)</td>
</tr>
<tr>
<td>Intention 50%</td>
<td>0.0727</td>
<td>0.0817***</td>
</tr>
<tr>
<td></td>
<td>(0.117)</td>
<td>(0.0261)</td>
</tr>
<tr>
<td>Intention 80%</td>
<td>-0.0715</td>
<td>0.304***</td>
</tr>
<tr>
<td></td>
<td>(0.105)</td>
<td>(0.0283)</td>
</tr>
<tr>
<td>Intention 100%</td>
<td>-0.376***</td>
<td>0.355***</td>
</tr>
<tr>
<td></td>
<td>(0.101)</td>
<td>(0.0318)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.964***</td>
<td>0.544***</td>
</tr>
<tr>
<td></td>
<td>(0.0964)</td>
<td>(0.0184)</td>
</tr>
</tbody>
</table>

**Observations**: 1,053 | 3,895
**Number of names**: 47 | 49
**R-squared**: 0.071 | 0.052

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1
11. Supplementary Material

11.1 Instructions

“Coordination Game: Instructions for Player B

You are participating in a Coordination Game. As a show up fee you have already earned 3 euros. Additionally you could earn up to 20 euros more in this game depending on your decisions and those of other participants. Pay close attention to the instructions.

The dilemma in this game is the following:

Two travelers come home from a remote island where they bought identical souvenirs. At their arrival the passengers find out their souvenirs have been destroyed. The airline will have to pay a compensation equal to the price of the souvenir to the passengers.

For the airline it is impossible to know the exact price of the souvenir, nonetheless the airline has estimated an approximate range in which the price may fall.

On the search for the honest price of the souvenir the airline separates the 2 travelers so they cannot confabulate, and asks them to report separately the price of the object.

If both report the same price it would be reasonable to assume that both are saying the truth and the airline will reimburse this quantity of money to each of the passengers.

If the passengers report different prices, the airline will take the lowest price as the true cost of the souvenir and each passenger will receive this amount as compensation. Nevertheless, a discount on the amount of the compensation is going to be made to the passenger that reported the Highest Price as a Penalty. This amount will be given to the passenger that reported the Lowest Price as a Reward for his “honesty”.

In this coordination game there are two types of Passengers: A and B. Each Passenger A will be randomly matched to play with a Passenger B.

Each round two different quantities of money that the airline is proposing to the passengers as compensation for the souvenir will be shown, as well as the amount of money the airline will impose as reward/penalty if two different prices of the souvenir are reported.

Each participant will take the decision separately on which quantity to report as the price of the souvenir. This procedure will be repeated for 100 rounds.

At the end of the game, one of the rounds will be randomly selected to be paid. The payment you will receive will depend on your decision and the one made by the other participant in the round chosen.

Example of Decision for one round

The airline offers the following amounts as compensation to the passengers:
(The amount of the reward/penalty appears in the center of the screen in red)

![9€ 5€ 6€](image)

- If both passengers report that the souvenirs price was 9 euros, both receive 9 euros
- If both passengers report that the souvenirs price was 6 euros, both receive 6 euros
- If different prices are reported, both passengers will get the Lowest Price and additionally:
  - The passenger that reported the Lowest Price receives a Reward
Low Price (6 euros) + Reward (5 euros) = Receives 11 euros

- The passenger that reported the **Highest Price pays a Penalty**

Low Price (6 euros) - Penalty (5 euros) = Receives 1 euro

Passengers A and Passengers B will take decisions in a different way.

Passengers A will define their strategy for a block of 20 rounds and Passengers B will define their strategy round by round.

**You are reading the Instructions for Passenger B.**

Passengers A will be asked to define a strategy for a block of 20 rounds by typing in the decision screen the number of rounds out of 20 in which they want to choose the High Price. Passenger A will have to determine this decision for 20 rounds with no prior information. The computer will randomly assign one decision from the reported strategy to each round, and after 20 rounds they will be able to update their strategy.

Before going to the decision screen were you select the price you want to choose, you will be shown for 2 seconds the strategy reported by the Passenger A you were matched with to play the round. The strategy shows the number of times the Passenger A decided to pick the High Price of the souvenir in a green bar in the following way:

![Examples of how you will see Passengers A reported strategies](image)

Each round you will be randomly matched with the Passenger A in the room and you will have 2 seconds to see his strategy.

Once you have seen Passenger A's strategy, you will be shown another screen with the two prices offered by the airline as well as the amount of the penalty. You will then select one of the two prices.

![Examples of decision screen](image)

One of the 100 rounds will be randomly selected to be paid at the end of the game. Your earnings and those of the Passenger A assigned to play with you that round will depend on both your decisions.

**CONTROL QUESTIONS” “**