Erasmus School of Economics

Erasmus University Rotterdam

Prior Predictions Influence Social Preferences

Behavioural Economics Master's Thesis

Patrick Sharpe

504670

Supervised by Kirsten I. M. Rohde

Second Assessor:

Date of Final Version:

Abstract

Studies in the domains of strategy and risk preferences observe that subjects make more rational choices when they are first asked to predict the behaviour of others. However, less is known about the impact of prior predictions on social preferences. This study uses an experiment to discern whether 'prediction then choice' changes how subjects approach social preference dilemmas. In particular, the expectation is that choices become more pro-social after subjects either predict others' actions or beliefs. The results confirm that prior predictions influence social preferences. Notably, subjects make more selfish choices when they are first asked to predict the action someone else would choose.

Acknowledgments

I would first like to thank my thesis supervisor Professor Rohde for her patience, expertise, and enthusiasm in guiding me through this project. I will not forget our entertaining conversations, both in person and oceans apart. I would also like to thank Professor Iriberri for her helpful advice concerning her method for eliciting preference types. Finally, I would like to thank my family and friends for completing the survey, especially those who went the extra mile by forwarding it to others.

Contents

1 Introduction

Many of us would be delighted to win £50 in the lottery, but less so if we discovered our neighbour won the jackpot. Our responses to the fortunes of others in relation to ourselves are known as social preferences. Potential applications of research into social preferences range from increasing charitable donations to investigating tolerance of wealth inequality. This study seeks to add to the literature by examining the influence of 'prediction then choice' in an experimental setting. Typically, 'prediction then choice' involves someone making a prediction about someone else before making their own decision. Research into this sequence in the domains of strategy and risk preferences finds that prior predictions often result in more rational choices. However, less is known about the effect of prediction then choice in a social preferences setting. An investigation in the context of social preferences could inform potential uses for 'prediction then choice' as a device to influence the ways people act towards others and respond to differences in wealth.

Using an experiment, this study compares the influence of three types of prediction on choices. Before making choices, subjects are either asked to predict what action another subject would choose (empirical treatment), what action others would say one should choose (normative treatment), or what action they would have chosen one hour ago (baseline). The choices subjects make are analysed independently as well as used to map subjects to four distinct social preference types. These types consist of selfish, social welfare maximising, inequity averse and competitive.

The results show that the type of prior prediction a subject makes does have an influence subsequent choices. The main finding is that subjects in the empirical treatment make more selfish choices than subjects in the baseline group. Further, subjects in neither treatment group produce actions or preference types that are more pro-social than the baseline group. Unexpectedly, this contradicts the findings of Krupka and Weber (2009) who, using similar treatments, find that subjects make more pro-social choices after first predicting the actions or beliefs of someone else.

The paper is organised as follows: Section 2 is a literature review of social preferences and the current research into prediction then choice in experimental settings. Section 3 explains the experimental procedure, treatment groups, and method for classifying subjects by their 'most compatible' preference types. Section 4 introduces general hypotheses, as well as hypotheses relating to four theories outlined in the literature review. Section 5 provides descriptive

statistics of the experiment data. Section 6 provides the results of the hypotheses tests and Section 7 discusses the key findings. Finally, Section 8 offers some concluding remarks.

2 Literature Review

This section provides a more comprehensive overview of social preferences and their wider applications. Additionally, I review the existing research into prediction then choice in the context of social preferences and other experimental settings. Finally, I outline four theories which concern why predictions are found to influence choices.

2.1 Social Preferences and Types

Traditional economics allows for many forms of diverse preferences. We can, for example, model different preferences for the goods we consume or the risks we take. However, traditional economics assumes that all agents are exclusively motivated by their own material interests. This assumption would appear to be at odds with the provisions made for diversity in so many other forms of preferences. Further, the validity of this assumption is challenged by more than two decades worth of experimental findings. This literature explores the nature of social preferences by examining how agents react when material payoffs for themselves and others are varied in a non-strategic context (see, for example: Levine, 1998; Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000; Andreoni & Miller, 2002; Charness & Rabin, 2002, Engelmann & Strobel, 2004). General themes in experimental findings are that sizeable proportions of agents are concerned with the outcomes of others, and the circumstances under which agents display concern are both varied and subjective. A classic example is the Dictator Game which involves a 'dictator' freely allocating a surplus between himself and someone else. Contrary to the assumption of selfishness, a meta-analysis finds that over 60% of dictators voluntarily display pro-social tendencies by choosing not to take the entire surplus (Engel, 2011).

Findings of this nature have informed the construction of social preference models – that is to say, utility functions including parameters which assign weights to others' payoffs. The arrangements of these parameters can represent different preference types, with each arrangement implying the conditions where weights are assigned to another's payoff. The ability to distinguish between preference types helps us to make sense of systematic behaviour. Charness and Rabin (2002) (C&R hereafter) introduce a model that explores conditions where an agent's payoff is either ahead (they earn more) or behind (they earn less) someone else's payoff. This allows for the modelling of four distinct types of preferences: selfish, social-welfare maximising, inequity averse, and competitive.

Selfish types

These agents place no weight on others' outcomes. Perfectly compatible with traditional assumptions, they always look to maximise personal payoff.

Social-welfare maximising types

These agents prefer efficiency and place a positive weight on others' outcomes. At some level, they are prepared to trade-off personal earnings to increase the aggregate payoff.

Inequity averse types

These agents experience a disutility from inequitable distributions of material resources. They place a positive weighting on another's outcome when this person earns less than a reference point, and a negative weighting when this person earns more than a reference point. Usually, this 'reference point' is modelled as the agent's own earnings. This means these agents experience disutility from differences between their own payoff and the payoffs of others. Consequently, at some level, these agents will tradeoff personal earnings to narrow the inequity.

Competitive types

These agents place a negative weight on others' outcomes. This means, both when the agent earns more or less than someone else, they are prepared to sacrifice personal earnings to reduce the earnings of someone else by a larger amount.

Some departures from C&R's standard modelling conditions have allowed for additional preference types to be identified, such as Fisman, Kariv and Markovits (2007) who distinguish a 'lexself' type[1] from selfish, and Engelmann and Strobel (2004) who distinguish a 'maximin'[2] type as a subset of the inequity averse type. However, using the standard C&R model as a base, several studies have investigated actions and preference types beyond the two dimensional selfish-altruistic divide. Cabrales, Miniaci, Piovesan, and Ponti (2010) explore the relationship between social preferences and the contracts that workers are offered and choose. In a lab experiment, the authors estimate subjects' parameter values under C&R's modelling

[1] Lexself or lexicographic preferences are when an agent prioritises their own payoff but will maximise the payoffs of others when their own personal payoff is unaffected by their decision.

[2] Maximin (Rawlsian) types prefer options which provide the greatest payoff to the least well-off in the group.

framework and compare these to the results of a simulated employment game. They find that these parameter estimates account well for workers' and employers' observed contract choices, and workers are more likely to choose a contract offered by an employer with more similar social preferences to their own. Through a similar method of estimating parameter values, Blanco, Engelmann, and Normann (2011) conclude that inequity aversion observed at the aggregate level is not found when conducting a within-subject analysis[3]. Iriberri and Rey-Biel (2013) also conduct a within-subject analysis but, rather than estimating parameters, choose to categorise subjects directly into the four distinct types described previously. Their analysis reveals substantial preference heterogeneity with 44% of their sample identified as selfish, 21% as social-welfare maximisers, 25% as inequality averse and 10% as competitive.

2.2 The Practical Relevance of Type Elicitation

'Lab-based' measures of social preferences are not always replicated in 'field-based' measures. Galizzi and Navarro-Martinez (2018) conduct a meta-analysis on studies comparing social preferences elicited in the lab and the field. They find that only 40% of the lab-field correlations and 38% of the lab-field regressions reported show significant associations.

Despite this, the authors defend lab-based social preference experiments by highlighting the value that they can provide through insights into important behavioural patterns. For instance, Cabrales et al.'s (2010) findings are one example of how understanding social preference types could lead to prospective employees being more efficiently matched with suitable contracts and work environments. In addition, Paternoster, Jaynes, and Wilson (2017) test the relationship between social preferences elicited in the lab and intentions to drink and drive. They find that those with strong other-regarding preferences were less likely to report that they would drive while drunk in a hypothetical scenario. Further, the threat of sanctions was more of a deterrent to selfish types. In another example, Kerschbamer, Sutter and Dulleck (2016) find that experimental markets for credence goods[4] do not function as efficiently as expected

[3] Blanco et al. (2011) base their analysis on the social preference type model introduced by Fehr and Schmidt (1999) which models inequity aversion. The C&R model extends this framework by allowing for a broader array of type distinctions.

[4] Goods where sellers have an asymmetric knowledge advantage of the optimal quality for consumers. An example is a car repair service where the mechanic knows more about the type of service the vehicle needs than the owner (Kerschbamer et al., 2016). The quality of some credence goods can be verified upon ex-post evaluation, such as the owner of a car noticing that fault reappears only shortly after repairs.

because less than a quarter of sellers act in accordance with the 'weak efficiency loving' preference type (lexself) assumed to be standard. As a result, Kerschbamer et al. (2016) highlight the potential importance of social preference type testing for helping employers identify those best suited to the roles they will perform.

2.3 Intervening in Preferences

Understanding the stability of social preferences is important for relating experiment-derived preference types to behaviour. Well-known literature on nudges (Thaler & Sunstein, 2009) and choice architecture (Johnson et al., 2012) shows preferences to be context dependent in domains such as healthy eating and pension contributions. However, comparatively little is known about the susceptibility of social preference types to influence.

One popular method for investigating preference stability in experimental settings is testing for order effects. Order effects concern the sequence in which agents process information and formulate beliefs (Hogarth & Einhorn, 1992). Under a theory of stable preferences, the order in which information is presented would have no effect on resulting behaviour. However, the behavioural literature proposes that, rather than calling upon a stable set of internal preferences, agents use mental procedures that allow them to generate preferences when called for. This should not be interpreted as individuals having no preferences; rather, individuals have strategies they use for assembling preferences and these are influenced by changes in contexts (Moore, 1999). Relatedly, 'query theory' (Johnson, Keinan & Häubl, 2007) proposes that individuals deconstruct decisions through asking series of introspective questions. Notably, order matters: the first query results in a richer and more heavily weighted representation than the second (Johnson et al., 2007).

Several experiments show that preferences may be influenced by first asking subjects to make a prediction of what someone else will do, and second, asking subjects to choose for themselves. This method of asking for a prediction then choice externalises the process of introspective questioning proposed by query theory. It ensures that subjects ask, 'what would someone else do?' and 'what should I do?' in a set order.

2.4 Experimental Evidence of Prediction Then Choice

Experiments in the domains of strategy and risk preferences find evidence that first predicting the actions of someone else improves the rationality of decision making in personal choices. In a strategic setting, Croson (2000) investigates whether incentivised prediction then choice

influences equilibrium play in a public goods game and prisoners' dilemma. In both cases, first eliciting predictions increased the frequency of choices consistent with rational expectations, resulting in more equilibrium play. According to Croson (2000), the most likely explanation for these results is that prediction then choice encourages subjects to think about the problem differently, and in particular, to think more like a game theorist by focussing on the dominant strategy. Similarly, in a social dilemma setting where the dominant strategy is to cooperate, Blanco, Engelmann, Koch, and Normann (2014) observe that cooperation increases when subjects first guess how many of the other 9 participants are co-operators. Relatedly, using a matching pennies game, Rutström and Wilcox (2009) find that first eliciting predictions has an effect in early rounds that is comparable to that of experience in later rounds.

However, some experiments in strategic contexts have not replicated the same results. Contrary to Croson (2000), Gächter and Renner (2010) find that prediction then choice in public goods games increases contributions. The authors propose that the contradictory findings could be the result of Croson (2000) asking subjects to predict the sum of others' contributions, whereas their experiment asks for the average, subsequently leading to different decisions. In a separate study, using a set of two-person normal-form games, Costa-Gomes and Weizsäcker (2008) find that there are no statistical differences in equilibrium play between groups which first predict and then choose, and groups which follow the reverse order.

In terms of risk preferences, findings appear to be in more agreement. Faro and Rottenstreich (2006) report that individuals predict others' choices to be closer to risk neutrality than those choices actually are. That is, risk seeking and risk averse agents predict others' risk preferences to move in the same direction as their own, but to a lesser extent. Their 'Experiment 3' also finds that subjects who made predictions then choices showed more risk neutrality in decisions for themselves. Arguing that risk neutrality is normatively desirable in many managerial and other contexts, the authors note the potential that prediction then choice has as a debiasing mechanism in risk contexts.

Li, Rohde and Wakker (2017) explore this further by using a clear test of rationality and fully incentivised predictions. By asking gain and loss framed risk questions, the authors compare the frequencies of preference reversals between subjects who predict then choose, and subjects who follow the reverse order. They find that prediction then choice does not improve risk neutrality or reduce preference reversals for choices in gains. However, prediction then choice does improve risk neutrality for choices in losses.

2.5 Prediction Then Choice in the Context of Social Preferences

Experimental tests of prediction then choice in the context of social preferences are limited. Currently, Krupka and Weber (2009) offer the clearest insight. They use a simple dictator game where dictators may choose between either a 'pro-social' or a 'selfish' choice. However, first, subjects either make a prediction of what others do, what others say one should do, or make no prediction. Krupka and Weber found that both prediction treatments resulted in increased pro-social behaviour in the dictators' own choices. These findings suggest that first making a prediction of what others do, or say one should do, will increase pro-social tendencies. However, the binary distinction of types provided by Krupka and Weber is less complete than the preference types agents exhibit in other social preferences experiments.

A second important contribution is the aforementioned experiment by Iriberri and Rey-Biel (2013) that distinguishes between selfish, social-welfare maximising, inequity averse and competitive types. In addition to eliciting types, the second part of their experiment asks subjects to predict other dictators' choices. A comparison between choices in the first stage and predictions in the second stage reveals that different preferences types give different predictions of how they expect other dictators to have chosen. Notably, 89% of selfish types believe others to be selfish, while social-welfare maximising, inequity averse and competitive types expect fewer others to be selfish types (55%, 64% and 61% respectively). Further, social-welfare maximisers predict that others choose surplus creating actions with the highest frequency (31%), while competitive types expect others to destroy surpluses with the highest frequency. Iriberri and Rey-Biel (2013) interpret these results as lending support for the false consensus effect such that dictators overestimate the likelihood that others choose the same decisions as themselves. However, by asking for choices first, these findings only provide an insight into the relationship between choices and predictions and not the causal effect of prediction then choice.

As such, a gap in the literature presents itself for an investigation into the causal effect of prior predictions on choice using an expanded set of social preference types. Utilising what already exists, bridging the gap is achievable through using a combination of Krupka and Weber's (2009) treatments with the preference type elicitation method used by Iriberri and Rey-Biel (2013). The following section expands on how this translates to an experimental framework. First, however, the remaining subsections present four theories on why prior predictions influence choices. These theories are used to construct hypotheses in Section 4.

2.6 Anchoring and Adjustment

When predicting an unknown (e.g. the number of sweets in a jar), the 'anchoring and adjustment heuristic' (Tversky & Kahneman, 1974) refers to starting with some initial value and adjusting to get a final answer. However, although this heuristic can make the process of prediction more straightforward, different initial (anchored) values can lead to different final answers. Further, the effect persists even when initial values are arbitrary, and incentives are provided for accurate predictions.

In one of Tversky and Kahneman's (1974) several demonstrations of this effect, subjects were split into groups and each group was shown a wheel being spun which produced a number between 1-100. Next, subjects were asked to estimate whether the percentage of African Nations in the United Nations was higher or lower than the number produced by the wheel. Finally, subjects were instructed to estimate the actual percentage by adjusting from the initial number given by the wheel. The median percentage estimates were markedly different between groups. For example, for groups which had been given 10 and 65 by the wheel, median estimates were 25 and 45 respectively.

The underlying psychological processes involved in the anchoring and adjustment heuristic are believed to be the product of two mechanisms. First, it is a mental shortcut to start from an initial value and adjust until reaching a plausible answer. However, the first plausible answer reached is typically too close to the starting point and, thus, adjustments are typically insufficient (Epley & Gilovich, 2006). The second mechanism understood to be at work is known as confirmatory hypothesis testing (Chapman & Johnson, 1999, 2002). Once a starting value has been anchored upon, a person tests the hypothesis that this anchored value is correct. Because hypotheses are tested through an attempt to confirm them, the mental search to answer this hypothesis leads to the person retrieving information that is disproportionately biased in favour of confirmation. Together, these processes of insufficient adjustment and confirmatory hypothesis testing produce estimates biased towards anchored values (Chaxel, 2014).

In the context of 'prediction then choice' the theory would imply that predictions are a self-generated anchor from which adjustment is insufficient. Therefore, anchoring and adjustment would forecast there to be a discernible degree of matching between predictions and choices. Accordingly, the answer given as a prediction should prove influential in the outcome of the final choice.

2.7 The False Consensus Effect

The false consensus effect is responsible for agents believing their own beliefs, attributes and behaviours are relatively common (Ross, Greene & House, 1977; Dawes, 1989). 'Relative', in this context, does not refer to an individual believing his features are more common than the true population value. Rather, 'relative' refers to how individuals with certain features will predict these features to be more commonly held by others than individuals who do not share these features. For example, both smokers and non-smokers could overestimate the proportion of smokers in a population, but smokers will tend to overestimate this proportion to a greater extent.

There are four overlapping theoretical perspectives which contribute to this effect (Marks & Miller, 1987). The 'selective exposure and cognitive availability' perspective suggests that people surround themselves with others with whom they share similar qualities and values. When asked to consider the characteristics or beliefs of others, the most available social information sample comes from those the agent is exposed to most often. In turn, this available information informs predictions of similarity.

Relatedly, the 'salience and focus of attention' view posits that the most available information to predict the dispositions of others is one's own perspective. By this approach, the false consensus effect is similar to the anchoring and adjustment heuristic; however, in this case, the agent's own perspective is the self-generated anchor which is adjusted to make a prediction for someone else.

The 'logical information processing' explanation instead focusses on the situation that the individual finds themselves in when making a prediction. In an experiment, all participants have the shared experience of answering the questions and so agents may imagine that this shared experience translates to similar choices. This requires the assumption that agents believe the shared experience affects themselves and others in the same way.

Finally, the 'motivation' perspective emphasizes how predicting similarity can provide value to the agent. For example, an individual who predicts that others are similar to him may validate the correctness of his own position and maintain self-esteem (Marks & Miller, 1987).

These theories overlap and are believed to vary in influence depending on the context of prediction. However, they all produce the common expectation that predictions for others will reflect own beliefs, attributes and behaviours. Importantly, choices would be unaffected by

prior predictions since answers given as predictions derive from what the agent would choose for themselves.

2.8 Construal Level Theory

Humans can think about themselves in the past and in the future; or may imagine themselves as someone different entirely. The more that 'psychological distance' abstracts our perspective away from our current selves, the higher the level of construal (Liberman, Trope & Wakslak, 2007; Trope & Liberman, 2010). Construal Level Theory proposes that the degree to which psychological distance changes has an impact on belief formation and decision making.

Asking a decision maker to consider someone else's beliefs or actions is one way of increasing psychological distance. This distance lessens influences of subjective contexts and emotions which, in turn, results in decision makers placing more weight on the invariant features of a choice, such as the payoff outcomes.

Evidence from the domains of strategy and risk indicates that first making a prediction about others fosters more normatively desirable behaviour in one's own choices[5]. This could be because, when making a prediction about others, the increase in psychological distance lessens the influence of subjective emotions and contextual factors. In theory, this reduction in subjective influences then carries through to posterior choice, resulting in more objectivity. Therefore, construal level theory would forecast the psychological distance incurred by prediction then choice to result in different decisions.

2.9 Norms and Focussing

Social norms are common understandings of appropriate behaviour and attitudes when interacting with others and the environment (Krupka & Weber, 2009). In experiments, actions in line with social norms, such as reciprocity and fairness, are frequently adhered to at the expense of immediate self-interest; even when acting selfishly is not punished. For this reason, models such as C&R's, which incorporate social preferences into decision making, are required to understand a vast array of findings.

However, the strength of the influence that a norm plays in a decision maker's consideration set is not necessarily constant. Individuals may regard norms as valuable without always complying with them. A theory developed by Cialdini, Reno and Kallgren (1990) suggests that

[5] Examples are detailed in Section 2.4

a norm's influence depends on how much attention is drawn to it. A key finding is that 'focus', defined as a heightened state of awareness, can result in more norm-compliant behaviour. Focus is activated when certain cues or types of behaviour are made salient through observation or more direct consideration. In a state of focus, ideas, concepts or behaviours related to the activation are more readily accessed in subsequent thought and decision making. In the context of prediction then choice, the act of predicting focusses the agent on the actions or beliefs of others. If the agent believes others will behave in a certain type of way, the agent could be inclined to replicate this in their own behaviour.

3 Experimental Design and Procedures

This section describes the experimental design used by Iriberri and Rey-Biel (2013) and the changes made for use in the present study. In addition, the treatment groups and preference type classification procedure are explained in detail.

3.1 Iriberri and Rey-Biel's (2013) Design

The first stage of the experiment used by Iriberri and Rey-Biel (2013) (I&R hereafter) is designed to categorise subjects into compatible preference types. First, subjects are assigned to either the role of Decider or Receiver and are anonymously partnered with someone else of the opposite role. Over the course of the experimental stage, Deciders are presented with 16 decision tables. Each of these tables displays three options for set payoff distributions for the Decider and Receiver. It was explained to participants that Deciders would have sole discretion over selecting options. Participants were informed that, at the end of the experiment, each Decider would have one choice randomly chosen to pay out for themselves and their randomly partnered Receiver.

Of the three payoff distributions that Deciders had to choose from, one action always offered a 1 unit higher payoff to the Decider (selfish (S)), one action always increased the Receiver's payoff by an amount greater than 1 (surplus creating (C)), and one action always reduced the Receiver's payoff by an amount greater than 1 (surplus destroying (D)). For each table, the amounts by which options C and D would create and destroy surplus were equal. This value, s, ranges between 2-7 and $\frac{1}{s}$ can be thought of as the price of creating or destroying surplus. This is illustrated in Figure 1. The tables differed in the values of s, the starting values of x and y, whether the Decider's payoffs were ahead or behind the Receiver's, and whether this would change based on the decisions made.

	S	C	D
Decider	x	$x-1$	$x-1$
Receiver	y	$y+s$	$y-s$

Figure 1. Decision table composition

Source: Iriberri and Rey-Biel, 2013

3.2 Experiment Redesign

For this study, I&R's experimental method was redesigned for an online survey using different treatment groups. The survey consisted of an introduction, experiment stage and demographics questions. A copy of the introduction and an example decision table is provided in The Appendix.

The introduction outlines the rules and conditions of the experiment. Subjects were informed that: participants have equal chances of being paid, personal information will be used in this experiment only, and chosen participants can be paid in a preferred currency. The instructions revealed that all participants would make decisions as a Decider and would be assigned an anonymous partner. One partnership would have one of their choices paid for real, with a coin flip deciding whose Decider choices would govern payments. Finally, before the experiment began, participants were informed that the experiment was about understanding preferences and there are no right or wrong answers. This was provided to reduce misconceptions about the experiment being in some way strategic.

Upon completion of the experiment section, participants were asked demographics questions. These asked for age, gender, primary occupation, level of education and whether a university level economics course had been completed. Finally, participants could leave an email address by which they could be contacted if chosen to be paid.

10 decision tables were selected from the original 16 used by I&R. This reduction was seen as necessary because of the increased risk of dropout expected from a more time-consuming survey. Original tables with the values s = 2, 4, 6, and 7 were retained whilst tables with the values s = 3 and 5 were not used. I envisaged that this would reduce survey completion times whilst retaining sufficient observations for each subject to determine preference types. The decision tables are provided in Figure 2. However, upon analysing the results of the experiment, it became apparent that Tables 6 and 10 were problematic. In these tables, the actions determine whether the Decider's payoff is ahead or behind the Receiver's payoff. This involves a different choice process to decision tables where the Decider's position is unchanged across all available actions. As a result, these tables were omitted from further analysis.

Table 1 (s = 2)	Option 1	Option 2	Option 3
Decider	€20	€19	€19
Receiver	€5	€7	€3

Table 2 (s = 2)	Option 1	Option 2	Option 3
Decider	€4	€4	€5
Receiver	€22	€18	€20

Table 3 (s = 4)	Option 1	Option 2	Option 3
Decider	€17	€16	€16
Receiver	€8	€12	€4

Table 4 (s = 4)	Option 1	Option 2	Option 3
Decider	€20	€19	€19
Receiver	€5	€1	€9

Table 5 (s = 4)	Option 1	Option 2	Option 3
Decider	€4	€5	€4
Receiver	€24	€20	€16

Table 6 (s = 6)	Option 1	Option 2	Option 3
Decider	€13	€14	€13
Receiver	€5	€11	€17

Table 7 (s = 6)	Option 1	Option 2	Option 3
Decider	€7	€7	€8
Receiver	€23	€11	€17

Table 8 (s = 7)	Option 1	Option 2	Option 3
Decider	€16	€16	€17
Receiver	€1	€15	€8

Table 9 (s = 7)	Option 1	Option 2	Option 3
Decider	€7	€7	€8
Receiver	€10	€24	€17

Table 10 (s = 7)	Option 1	Option 2	Option 3
Decider	€10	€10	€11
Receiver	€21	€7	€14

Figure 2. Decision Tables

Adapted source: Iriberri and Rey-Biel, 2013

There are several additional changes from the original design. First, all survey participants make predictions and choices as a Decider – as opposed to the clear role distinction used in I&R's experiment. This change was made possible by informing participants that they would be anonymously partnered with another subject and the one partnership paid for real would have a coin flipped to determine whose Decider choices would govern the final payment allocation. The reason for departing from I&R's distinct role separation was to gather twice the choice data. Second, this experiment departs from I&R by not incentivising correct predictions of others' actions or beliefs. Aside from financial limitations, not incentivising predictions is relevant for a research interest in prediction then choice as a debiasing mechanism in everyday

situations. Since it would be unusual for correct predictions of others' social preferences to be rewarded under most circumstances, not incentivising predictions has some relevance for wider applications. Another departure from I&R is that subjects in this experiment were informed that only one partnership would have one of their choices played for real, whereas, in I&R's experiment, all participants received some material payoff. Again, financial limitations played a role in this decision. The final noteworthy difference is that the decision tables used in this experiment featured euro currency symbols. These were included to make the monetary stakes of the game more salient, which might have otherwise been lost on participants given the low likelihood of choices being played for real.

3.3 Treatments

The tables were presented to participants in a randomised order, with each table containing a Part A and B. Upon starting the experiment, participants were randomly assigned to the 'baseline' group or either one of the 'empirical' or 'normative' treatment groups. Part A required participants to make a prediction from the table based on their respective group. These were as follows:

Empirical group: *First, consider that your partner also faces an identical decision as a Decider in Part B. Please predict the option you think your partner is most likely to choose as a Decider.*

Normative group: *First, please select the option that you believe others would say you should choose.*

Baseline group: *First, please select the option you think you would have chosen 1 hour ago.*

The empirical and normative treatments are based on those used by Krupka and Weber (2009) and are similar to those used by Bicchieri and Xiao (2009). An empirical expectation is the choice that one expects most others to follow. A normative expectation concerns the choice which the subject believes others would say he should choose. According to Krupka and Weber (2009), focussing on the predicted choices or beliefs of others draws a subject's attention to norms of behaviour, leading them to evaluate their own decisions with these norms in mind. These expectations can be the product of past observations of others' conformity to norms, consequences for deviating, and projections of personal beliefs onto others (Bicchieri & Xiao, 2009).

The key difference between the empirical and normative prediction treatments concerns how directly norms are considered. Krupka and Weber (2009) suggest that the normative prediction

treatment focusses attention directly on pro-social norms of sharing, while the empirical prediction treatment is likely to produce a similar effect on behaviour only to the extent that thinking about what others do leads one to think about they should do. In Krupka and Weber's experiment where subjects have a binary choice between a selfish and pro-social action, both treatments had a comparable effect of significantly increasing the frequency of pro-social actions.

Relatedly, Bicchieri and Xiao (2009) find that the empirical predictions that subjects make for others are strongly related to what subjects choose for themselves. However, normative predictions are only a good predictor of what the subject chooses for themselves when normative predictions are aligned with empirical predictions. As such, Bicchieri and Xiao (2009) cast doubt upon whether normative beliefs have any significant effects on personal choices. Therefore, it is important to investigate both treatments to determine whether there is a difference in effect.

The expectation for the baseline prediction is that subjects would, in most cases, have chosen the same one 1 ago as they choose in the moment. This prediction was included so that those in the baseline group answered the same total number of experiment questions as those in the other groups, as well as reading the same experimental instructions. In theory, this allows for the baseline group to be used for neutral comparisons with the empirical and normative prediction treatments.

3.4 Eliciting Preference Types

Using the data from the experiment, within-subject analyses can determine preference type compatibility. These types are representative of four parameter arrangements of C&R's model. In the context of this experiment, the form of C&R's model is shown in Equation 1:

$$u_D(\pi_{Rta}, \pi_{Dta}) = (\rho r + \sigma v)\pi_{Rta} + (1 - \rho r - \sigma v)\pi_{Dta} \quad (1)$$

For $t = 1, \dots T$ and $a = \{S, C, D\}$

Where:

$r = 1$ if $\pi_{Dta} \geq \pi_{Rta}$

$v = 1$ if $\pi_{Dta} < \pi_{Rta}$

The function describes how a Dictator's utility (u_D) is a weighted sum of his own and the Receiver's material payoffs, with weightings dependent on whether the Dictator's payoff is ahead or behind the Receiver's. The function shows the utility for a decision table (t) with payoffs the product of an action (a). π_{Dta} and π_{Rta} are the respective payoffs for Deciders and Receivers based on a certain decision table and action. The parameters ρ and σ are the weights given to the Receiver's payoff when the Dictator is ahead or behind respectively.

Distinct preference types are given by four arrangements of ρ and σ in C&R's model. These four arrangements denote selfish (SF), social welfare maximising (SW), inequity averse (IA) and competitive (CP) types. In the context of this experiment, each preference type informs a set of behavioural rules for type compatibility. The preference type that a subject is most compatible with is determined by the fewest violations of the behavioural rules associated with that type[6]. These behavioural rules and parameter arrangements are detailed as follows:

SF types always choose the selfish action (S). Surplus creating actions (C) and surplus destroying actions (D) are recorded as errors. This is given by the arrangement $\rho = \sigma = 0$ in C&R's model.

For the remaining three preference types some further notation is needed. The remaining types all contain other-regarding aspects which means, at some value of s, subjects in these preference types will switch from choosing S to choosing either C or D for tables with the same or greater value of s. This is because increases in s equate to a decrease in the price of creating or destroying surplus. s_t denotes the value of s for the current decision table and s' denotes the switching value at which point and thereafter ($s_t \geq s'$) a subject decides to choose either C or D (as appropriate to the rule) rather than S. The value of s' is given by the lowest value of s at which a subject deviates from S in accordance with the behavioural rules of a type. For example, the SW type predicts that subjects will switch from S to C at s'. s' is considered the lowest value of s at which a C choice is made. For $s_t < s'$ the behavioural rule is for S to be chosen, whereas, for $s_t \geq s'$ the behavioural rule is for C to be chosen.

Further, C&R's model makes a distinction between decision tables where the Dictator's payoff is either ahead or behind the Receiver's. Ahead being when the Decider earns more than the Receiver, and behind being when the Decider earns less. This means there can be different switching points for ahead (s'_A) and behind (s'_B). Subscripts A and B indicate whether the

[6] It is possible for subjects to be most compatible with more than one type. This would be the case if a subject commits the fewest number of errors with respect to more than one set of behavioural rules.

Decider's payoff is ahead or behind the Receiver's respectively. Note, s_A' and s_B' need not be equal.

When ahead, SW types choose S when $s_{At} < s_A'$ and C when $s_{At} \geq s_A'$. Similarly, when behind, SW types choose S when $s_{Bt} < s_B'$ and C when $s_{Bt} \geq s_B'$. Once a C choice establishes a value for s_A', choices for tables with $s_{At} \geq s_A'$ must also be C or else an error is recorded. The same is true for choices made when the Dictator is behind. D choices are always recorded as errors. This gives SW parameter arrangements under C&R's model as: $0 < \sigma \leq \rho \leq 1$ and $0 < \rho \leq \sigma \leq 1$.

Further, I&R note a requirement for SW types that there is at least one incident of switching from S to C when the Decider's payoff is behind so that there is differentiation between SF and SW types. Conversely, subjects who choose C when ahead but only S when behind are classified as IA since their behaviour always yields a more equalitarian distribution of payoffs (Iriberri and Rey-Biel, 2013). As such, another parameter arrangement for the SW type is $0 = \rho < \sigma \leq 1$.

When ahead, IA types choose S for $s_{At} < s_A'$ and C for $s_{At} \geq s_A'$. When behind, IA types choose S for $s_{Bt} < s_B'$ and D for $s_{Bt} \geq s_B'$. As such, s_A' is given by the lowest value of s_A for which C is chosen and s_B' is given by the lowest value of s_B for which D is chosen. It is a violation of the IA behavioural rules to choose S when $s_{At} \geq s_A'$ or $s_{Bt} \geq s_B'$. Further, errors are incurred by any choice of C when behind and any choice of D when ahead. I&R also include that subjects who select from only S and D actions are CP rather than IA. This gives the IA type parameter specification as $\sigma \leq 0 < \rho \leq 1$.

When ahead, CP types choose S for $s_{At} < s_A'$ and D for $s_{At} \geq s_A'$. Similarly, when behind, CP types choose S for $s_{Bt} < s_B'$ and D for $s_{Bt} \geq s_B'$. This means s_A' and s_B' are given by D choices at the lowest values of s ahead and behind respectively. Similar to SW and IA, choosing S when $s_{At} \geq s_A'$ or $s_{Bt} \geq s_B'$ is a violation of the CP type behavioural rules. Additionally, any C choice is recorded as an error. Also note that, to differentiate from SF types, there must be at least one recorded switch from S to D either ahead or behind. This gives the CP parameter specifications as $\sigma < \rho \leq 0$ and $\rho < \sigma \leq 0$ and $\sigma = \rho < 0$.

To reiterate, under these four sets of behavioural rules, errors are recorded under two conditions. First, when a choice is selected that is categorically inconsistent with the preference type; for example, an SF type choosing C or D. Second, when a subject chooses S when $s_{At} \geq$

s'_A or $s_{Bt} \geq s'_B$. Each instance of a rule violation is recorded as an error. Individuals are considered 'most compatible' with the preference type or types for which they make the fewest number of errors.

There are some potential parameter arrangements that are not covered by any of the four types, such as $\rho < 0 < \sigma \leq 1$ which would give the unusual prediction that an agent is prepared to destroy surplus when ahead but create surplus when behind. The four preference types are not intended to cover the complete range of all behaviour. Rather, these four types are expected to account for the most prevalent systematic preferences, each distinguished by comprehensible motivations.

4 Hypotheses

This section presents general hypotheses, as well as hypotheses relating to the four theories presented in Section 2.

4.1 General Hypotheses

This study predicts that the type of prior prediction made has an influence on a subject's subsequent choices. This could be reflected in either chosen actions or preference types differing between groups. Therefore, some general hypotheses are as follows.

H1: Chosen actions differ between groups.

H2: Chosen preference types differ between groups.

4.2 Anchoring and Adjustment

Whereas most tests for anchoring and adjustment involve estimating continuous values, this experiment uses nominal choices. However, it is reasonable to believe that the heuristic remains relevant in the present context. In a related study of risk preferences, Li, Rhode and Wakker (2017) use simple yes/no questions to investigate the effect of subjects first predicting someone else's choice and then choosing for themselves. They found evidence of a relationship between prediction and posterior choice supporting a process of anchoring and adjustment. In this sense, predictions are a self-generated anchor and adjustment is insufficient so that predictions and posterior choice are matching.

For the present study, it is plausible that predictions are an anchor used in subsequent choice. Therefore, at the very least, anchoring and adjustment cannot be ruled out when there is matching between prediction and choice answers. However, for evidence that anchored predictions are a dominant factor in choice decisions, it would also be expected that rates of matching between predictions and choices are equivalent across prediction groups. This would imply that anchoring and adjustment is equally as strong regardless of the type of prediction made. Meeting this criterion would replicate the finding that anchors are a dominant factor in decision making even when the relatedness of predictions to choices is varied. As such, weak and strong forms of the anchoring and adjustment hypothesis are as follows:

AA1: The rate of matching between predictions and choices is stronger than chance.

AA2: The rate of matching between predictions and choices is stronger than chance, and rates do not differ between groups.

4.3 The False Consensus Effect

In this experiment, the false consensus effect hypothesis is similar to anchoring and adjustment to the extent that predictions will match choices. However, since predictions are grounded on what subjects plan to choose for themselves, the treatment processes should have no effect on choice. As such, the false consensus effect predicts that predictions and choices will match, but there will be no significant differences between the choices made by any groups. Therefore, the false consensus effect hypothesis is as follows:

FC: The rate of matching between predictions and choices is stronger than chance, and there are no differences between the choices made by groups.

4.4 Construal Level Theory

In this experiment, construal level theory would propose that subjects in the empirical and normative prediction treatments experience psychological distance when they first consider the actions or beliefs of others. Conversely, psychological distance is predicted to be negligible for subjects in the baseline group who only consider what they themselves would have chosen an hour ago.

However, unlike strategy and risk, there is no agreement over which social preferences are normatively desirable. Therefore, using construal level theory, it is difficult to predict which direction psychological distance will move social preferences, if at all. Although, this does not mean that construal level theory can be overlooked in the context of social preferences. Any evidence of systematic differences in the preference types of subjects undertaking the treatments compared to the baseline could be explained by the increase in psychological distance. Therefore, the construal level theory hypotheses are as follows:

CLT1: Treatment group choices are different from baseline group choices.

CLT2: Treatment group preference types are different from baseline group preference types.

4.5 Norms and Focussing

The norms and focussing theory produces three sets of hypotheses. First, if the degree of matching between predictions and choices is greater in the empirical group than the normative

group, this could indicate that first predicting how someone else acts focusses a subject on the norm of reciprocity. For example, an agent would respond to anticipated selfishness with selfishness, or anticipated kindness with kindness. Like anchoring and adjustment, this provides the expectation that predictions will be replicated in choices. Further, we might expect this matching to be strongest when agents are focussed directly on predicted actions rather than beliefs. Therefore, higher rates of matching between predictions and choices for the empirical group could be indicative of subjects reciprocating what they anticipate their partner to do. The reciprocity hypothesis is as follows:

N&F1: Rates of matching between predictions and choices are higher in the empirical treatment than the normative treatment.

Second, previous evidence also suggests that heightened focus on norms will increase pro-social choices. In a similar setting, Krupka and Weber (2009) find that participants show a greater tendency to behave pro-socially after making predictions in both treatments. Therefore, in this experiment, it would be expected that the treatments also focus subjects on pro-social behaviour. Unlike construal level theory, this interpretation of the norms and focussing theory allows for a hypothesis with directional predictions: participants' choices in the treatment groups will be more pro-social compared with the baseline. In the treatment groups, this will be represented by a greater proportion of C choices and SW or IA preference types in Part B. Note, this is irrespective of what participants' predictions are since, for example, even if a participant imagines others to be selfish, the act of considering someone else may activate a focus which leads to more pro-social personal choices. These hypotheses are as follows:

N&F2: There are a higher proportion of C choices in the treatment groups than the baseline group.

N&F3: There are a higher proportion of SW and IA compatible preference types in the treatment groups than the baseline group.

Third, Krupka and Weber (2009) also argue that the normative treatment draws attention more directly to the norms of behaviour. Therefore, we might also expect a higher frequency of pro-social choices and types in the normative group compared with the empirical group.

N&F4: There are a higher proportion of C choices in the normative treatment than the empirical treatment.

N&F5: There are a higher proportion of SW and IA compatible preference types in the normative treatment than the empirical treatment.

5 Descriptive Statistics

The survey was released in June 2019. The distribution channel consisted of my personal and extended network as well as participants from the reciprocity-based survey completion website SurveySwap. In total, 397 people opened the survey with 272 users submitting complete responses – a completion rate of 68.5%. The median time taken to complete the survey was 7 minutes 32 seconds. 5 participants were omitted for completing the survey in under 3 minutes[7].

Overall, the sample is 48% male, with an average age of 30. The youngest participant is 18 and the oldest is 85. 57% of participants are students and 34% are in employment. In terms of education, 39% said they had completed a university level course in economics and only 3 subjects had not completed high school education. At the end of the experiment, 81% chose to leave their email address for the chance to be paid. Decision Table 6 (see Figure 2) was chosen to be played out for real with one Decider earning €14 and one Receiver earning €11.

Of the 267 complete responses, 81 participants were in the baseline group, 91 were in the empirical treatment and 95 were in the normative treatment. This produced 2136 choices and the same number of predictions[8]. Table 1 displays a summary of demographics and checks for balance between the three groups. Differences are tested for significance using the Mann-Whitney U test for age distributions, and two-tailed Fisher exact tests for binary variables. The tests show no significant differences at the 10% level and indicate that subjects are well-distributed between groups.

[7] Completion times beneath this cut-off are deemed too short for the subject to have given the experiment sufficient attention.

[8] This excludes decision tables 6 and 10 in Figure 2.

Table 1 - Demographics Summary and Balance Tests

					Difference		
	Baseline	Empirical	Normative	Overall	Baseline vs. Others	Empirical vs. Others	Normative vs. Others
Indicator	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Average Age (years)	29.78	30.45	30.73	30.34	-0.81	0.16	0.59
Prop. MALE	0.49	0.51	0.43	0.48	0.03	0.05	-0.07
Prop. STDNT	0.53	0.63	0.55	0.57	-0.06	0.09	-0.03
Prop. JOB	0.38	0.27	0.37	0.34	0.06	-0.10	0.04
Prop. HEDUC	0.81	0.80	0.73	0.78	0.05	0.04	-0.08
Prop. ECON	0.35	0.42	0.41	0.39	-0.07	0.04	0.03

Notes: Columns 1-4 present a summary of demographics overall and for each of the treatment groups. 'Average Age' shows the mean age of each sample in years. 'Prop.' Indicates that binary variables are given as proportions. 'MALE' = 1 if a subject is male. 'STDNT' = 1 if a subject's primary occupation is studying. 'JOB' = 1 if a subject's primary occupation is part-time or full-time employment. 'HEDUC' = 1 if a subject has completed some form of higher education. 'ECON' = 1 if a subject has completed a university level economics course. Columns 5-7 report differences between a selected group and the other groups based on demographic indicators. The values provide the difference in means or proportions. The differences between age distributions are tested for significance using the Mann-Whitney U test. The differences between binary variables are tested for significance using two-tailed Fisher exact tests. Asterisks indicate the level at which differences are found to be significant.

*** Significant at the 1% level

** Significant at the 5% level

* Significant at the 10% level

Table 2 displays the overall prediction and choice proportions of actions and preference types. Across all three groups, C was chosen the most frequently, followed by S then D. In terms of predictions, orders differ by group. In the baseline, the C, S, D ordering holds for both predications and choices. However, in both the empirical treatment and the normative treatments, subjects predict others will choose S and C actions at more comparable rates.

In terms of preferences types, Table 2 displays the percentages of subjects whose actions result in the specified type being recorded as their 'most compatible' (or amongst their 'most compatible' if the subject commits the fewest amount of errors for more than one preference type). For choices, the ordering SW, IA, SF, CP is consistent across all groups. In terms of predictions, this ordering holds for the baseline and empirical groups; however, for the normative group the order of prediction preference types is SW, SF, IA, CP.

Table 2 - Overall Proportions of Action Types and Most Compatible Preference Types by Group

		Action			Preference Type (%)			
		S	C	D	SF	SW	IA	CP
Group		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Baseline	Prediction	0.38	0.56	0.06	13.6	59.3	30.9	11.1
	Choice	0.31	0.63	0.06	8.6	67.9	29.6	4.9
Empirical	Prediction	0.46	0.48	0.05	17.6	57.1	29.7	12.1
	Choice	0.39	0.55	0.06	11.0	58.2	34.1	5.5
Normative	Prediction	0.48	0.45	0.07	28.4	42.1	25.3	8.4
	Choice	0.34	0.61	0.05	16.8	63.2	24.2	4.2

Notes: Columns 1-3 display the proportions of S, C and D actions, rounded to two decimal places. Columns 4-7 display the percentages of individuals for which the specified type is their most compatible type (or amongst their most compatible types if the individual commits the fewest amount of errors for more than one type). This is shown separately for SF, SW, IA and CP types, rounded to two decimal places.

Differences between predictions and choices at the group level are also interesting. Table 3, shows the differences in proportions between predictions and choices at the group level. These differences are tested for significance using two-tailed Fisher exact tests. For all groups, significant differences exist between predicted and chosen proportions of S and C actions. Although, no significant differences are observed for any group in the proportions of predicted and chosen D actions. In addition, one-tailed Fisher exact tests confirm the directions of all the significant S and C differences at the 1% level. For all groups, S actions are predicted more than they are played and C actions are predicted less than they are played. In particular, the most pronounced differences between predictions and choices exist for the normative group. This indicates that subjects believe others would say they should choose selfish options at considerably higher rates than the subjects actually choose for themselves.

Table 3 - Proportions of Predicted Actions, Chosen Actions, and Differences

	Difference		
	Baseline Prediction vs. Choice	Empirical Prediction vs. Choice	Normative Prediction vs. Choice
Action type	(1)	(2)	(3)
S	0.06 **	0.08 ***	0.14 ***
C	-0.07 ***	-0.07 ***	-0.16 ***
D	0.01	-0.01	0.02

Notes: Columns 1-3 show differences in proportions between predictions and choices for the specified group. Differences are tested for significance using two-tailed Fisher exact tests. The null hypothesis for each test is that the difference is equal to zero. Asterisks indicate the level of significance at which the null is rejected.

*** Significant at the 1% level
** Significant at the 5% level
* Significant at the 10% level

6 Results

In this section, I test the hypotheses proposed in Section 4.

6.1 General Hypotheses

H1 is that chosen actions differ between groups. Overall distributions of chosen actions can be compared using Fisher exact tests. The null hypothesis for each test is that observed distributions are equal between the two compared groups. Comparing the baseline and empirical groups, the test rejects the null hypothesis at the 5% level. However, comparing the normative treatment to both the baseline group and the empirical treatment, the tests fail to reject the null at the 10% level. These results support H1 for the empirical treatment differing from the baseline. Although, further evidence is needed to determine whether the normative treatment produces discernibly different actions to the baseline and empirical groups.

H1 can also be tested by independently comparing proportions of S, C and D actions across groups. Independent comparisons help to pinpoint which actions are chosen differently. This investigation is carried out using two-tailed Fisher exact tests. Table 4 shows proportions of chosen action types and differences across groups. The proportions of S and C choices selected by the empirical group differ from the baseline at the 1% level. This reiterates the finding that choices in the empirical group differ from the baseline. Further, the tests also show that the empirical group and normative group differ in the proportions of chosen S and C actions at the 10% and 5% levels respectively. Therefore, this provides some support for H1 for a difference in choices between the empirical and normative groups. Moreover, Table 4 shows these differences between the empirical treatment and the other groups predominantly concern the proportions of S and C choices.

However, as before, there are no significant differences found between the baseline group and normative treatment. For the normative treatment, this independent analysis of action types does not provide the further evidence needed to conclude that the normative treatment produces different choices to the baseline. Therefore, H1 can only be confirmed for differences between the empirical treatment and the baseline, and the empirical treatment and the normative treatment.

Table 4 - Proportions of Chosen Action Types and Differences Across Groups

	Proportion of choices			Difference		
Action type	Baseline (1)	Empirical (2)	Normative (3)	Baseline vs. Empirical (4)	Baseline vs. Normative (5)	Empirical vs. Normative (6)
S	0.32	0.39	0.34	-0.07 ***	-0.03	0.05 *
C	0.63	0.55	0.60	0.08 ***	0.03	-0.05 **
D	0.06	0.05	0.05	0.00	0.00	0.01

Notes: Columns 1-3 show the three groups by their compositions of S, C and D proportions, given to 2 decimal places. Columns 4-6 show the differences in proportions between the first listed group and the second for each action type. Differences are tested for significance using two-tailed Fisher exact tests. The null hypothesis for each test is that the difference is equal to zero. Asterisks indicate the level of significance at which the null hypothesis is rejected.

*** Significant at the 1% level

** Significant at the 5% level

* Significant at the 10% level

H2 is that chosen preference types differ between groups. First, I test for group level differences between overall distributions of 'most compatible' preference types. These differences can be tested for using Fisher exact tests. The null hypothesis for each test is that observed distributions are equal between the two compared groups. For all comparisons between groups, the tests fail to reject the null hypotheses. As such, for overall distributions of most compatible preference types, there is no support for H2.

Differences between groups for each independent preference type can also be analysed. Table 5 shows the percentages of subjects most compatible with each preference type and the differences between groups. Using two-tailed Fisher exact tests, these differences are tested for significance. There are no significant differences for independent types when comparing any of the three groups. This means H2 is rejected on the basis of a 'most compatible' preference type classification.

Table 5 - Proportions of Most Compatible Preference Types & Differences Between Groups

	Most compatible choice types (%)			Difference (%)		
Pref. type	Baseline (1)	Empirical (2)	Normative (3)	Baseline vs. Empirical (4)	Baseline vs. Normative (5)	Empirical vs. Normative (6)
SF	8.6	11.0	16.8	-2.3	-8.2	-5.9
SW	67.9	58.2	63.2	9.7	4.7	-4.9
IA	29.6	34.1	24.2	-4.4	5.4	9.9
CP	4.9	5.5	4.2	-0.6	0.7	1.3

Notes: Columns 1-3 show the three groups by their compositions of SF, SW, IA and CP proportions, given to 2 decimal places. Columns 4-6 show the differences in proportions between the first listed group and the second for each preference type. Differences are tested for significance using two-tailed Fisher exact tests. The null hypothesis for each test is that the difference is equal to zero. Asterisks indicate the level of significance at which the null hypothesis is rejected.

*** Significant at the 1% level

** Significant at the 5% level

* Significant at the 10% level

However, in addition to testing subjects' most compatible types, it is also interesting to determine 'how compatible' subjects are with each type. This second measure of preference type classification can be illustrated by plotting differences between groups based on cumulative proportions of errors per type. For each group, this entails recording the cumulative frequencies of subjects who commit fewer than 1 errors, 2 errors, 3 errors, and so on for each preference type. The results for all four types are displayed graphically in Figure 3.

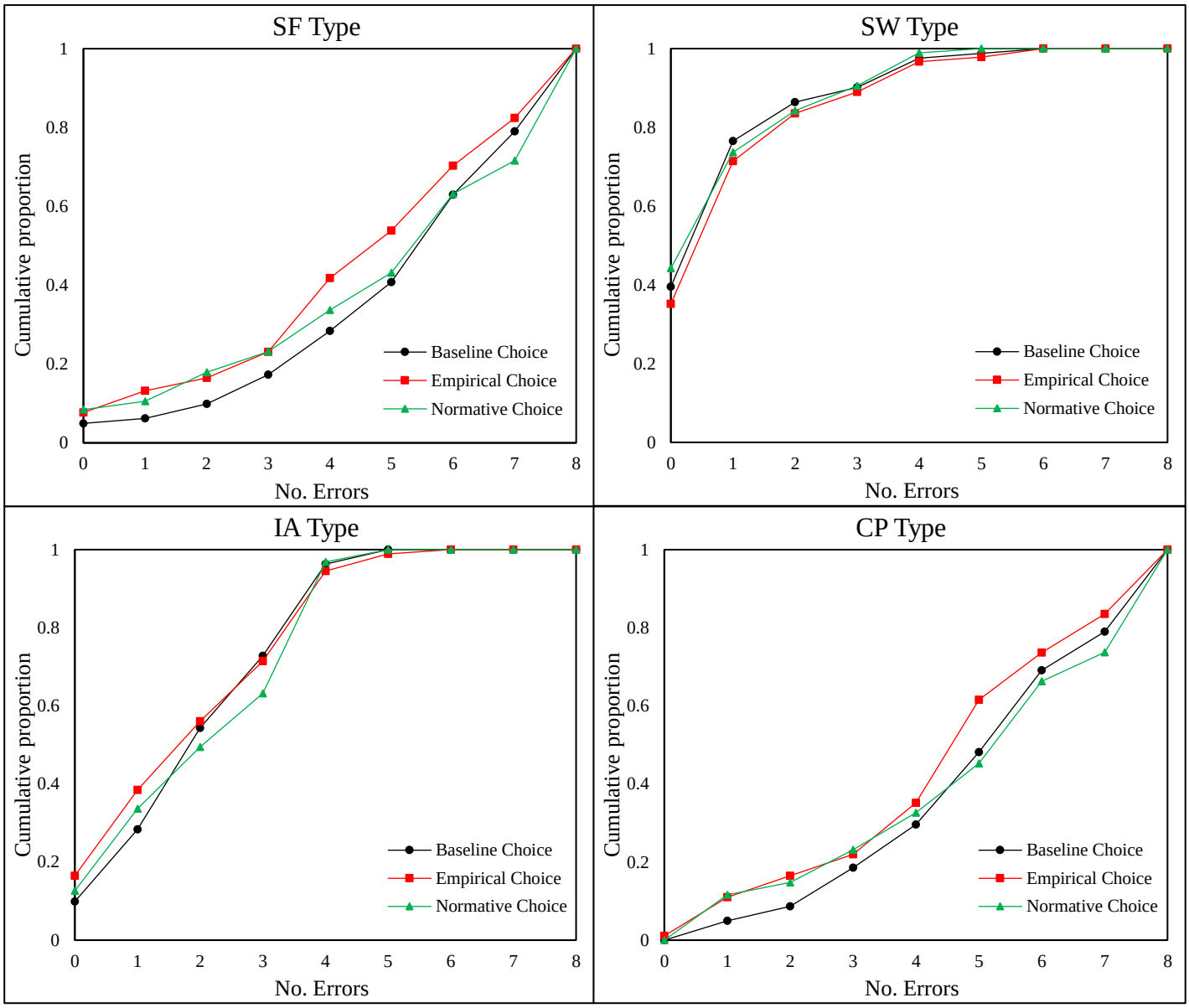

Figure 3. Cumulative error graphs

For the most part, there are no major discernible differences between groups. The exception is that the empirical group appears to show more type compatibility with the SF type than the baseline and normative groups. However, testing for significant differences in distribution using Kolmogorov Smirnov tests fails to reject the null hypothesis that the empirical group contains an equal amount of errors to the baseline and normative groups[9].

Therefore, in terms of H2, neither the 'most compatible' or 'how compatible' measures of preference types support the existence of differences between groups. As such, observations of

[9] No other differences in distributions were found to be significant.

significant differences in choices between groups are limited to the proportions of S, C and D actions.

6.2 Anchoring and Adjustment

The anchoring and adjustment hypothesis is that predictions match choices. This can be construed in weak and strong forms. The weak form, AA1, is that the rates of matching between S, C and D predictions and choices are stronger than chance. Comparatively, the strong form, AA2, adds the condition that rates of matching between predictions and choices do not differ between groups. Although a strict indicator, differences in rates of matching would suggest that the type of prediction made influences the extent to which predictions are anchored upon in subsequent choices. Further, when there are differences, it is impossible to confirm that subjects 'anchor' on their predictions. Rather, subjects could simply believe that the same answers are most appropriate for both their predictions and choices, without anchoring influencing their decisions.

Table 6 provides a within-subject summary of matching between predictions and choices. The values in the first row for Columns 1-3 can be interpreted as the rate of matching between predictions and choices for an average subject in each group. For example, the average subject in the baseline group gives 6.8 (out of a possible 8) matching answers for predictions and choices. The second row shows that this equates to a rate of matching of 85%.

First, using one-tailed Fisher exact tests, the rates of matching between predictions and choices are compared to the expected rates of chance (one third)[10]. For each test, the null hypothesis is that the rate of matching between predictions and choices is equal to or less than chance. For all groups, this is rejected at the 1% level. As a result, there is support for AA1 meaning, at the very least, anchoring and adjustment cannot be ruled out.

With regards to AA2, the null hypothesis is that the rates of matching between predicted actions and chosen actions are equal between the baseline and the treatment groups. Columns 4, 5 and 6 (Table 6) show that the null is rejected at the 1% level for all comparisons between groups. Therefore, AA2 is rejected. This means that the presence of anchoring and adjustment cannot be ruled out, however, it is apparent that its influence could depend on the type of prior prediction that is made.

[10] These tests are not shown on Table 6.

Table 6 - Individual Level Matching Between Predicted and Chosen Actions

				Difference		
	Baseline	Empirical	Normative	Baseline vs. Empirical	Baseline vs. Normative	Empirical vs. Normative
	(1)	(2)	(3)	(4)	(5)	(6)
Mean	6.8	6.0	5.1			
Proportion	0.85	0.75	0.64	0.10 ***	0.21 ***	0.11 ***

Notes: Columns 1-3 show the rates of matching between predicted actions and chosen actions for each group. The first row displays the mean number of matched predictions and choices per subject, given to one decimal place. The second row displays the mean rate of matching as a proportion of the maximum matching rate (up to 8 matched predictions and choices), given to two decimal places. Columns 4-6 show differences in proportions between the first specified group and the second. Differences are tested for significance using two-tailed Fisher exact tests. The null hypothesis for each test is that the difference is equal to zero. Asterisks indicate the level of significance at which the null is rejected.

*** Significant at the 1% level
** Significant at the 5% level
* Significant at the 10% level

6.3 False Consensus Effect

There are two parts to the FC hypothesis. First, predictions match choices and, second, there are no differences between the choices made by groups. There is some support for the first part of this hypothesis through the confirmation of AA1. However, to confirm the second part of the FC hypothesis, the proportions of S, C and D actions chosen must be shown not to differ between groups. As discussed previously when investigating the general hypotheses in Section 6.1, Table 4 shows that there are no significant differences between the baseline and the normative treatment regarding the proportions of S, C and D actions in choices. However, Columns 4 and 6 show that the empirical group differs from both the baseline and the normative treatment in terms of the proportions of S and C actions. Since choices significantly differ based on the type of prediction a subject makes first, this rejects the FC hypothesis.

6.4 Construal Level Theory

Construal level theory proposes that the psychological distance incurred by making predictions about others influences the choices subjects make for themselves. CLT1 is that treatment group choices differ from baseline group choices. This is straightforward to investigate by comparing the rates of S, C, and D choices between the treatment groups and the baseline. Table 4 shows that the choices made by the empirical group differ from the baseline, whereas the choices made by the normative group do not. For the empirical treatment, these findings support the

predictions of CLT1. It appears that first making a prediction of what someone else chooses subsequently influences the likelihoods of the subject playing S or C. More specifically, one-sample Fisher exact tests confirm that subjects in the empirical prediction group choose S more frequently and C less frequently than the baseline. However, CLT1 is rejected for the normative treatment.

CLT2 is that proportions of preference types differ between the treatment and baseline groups. Table 5 displays proportions of 'most compatible' preference types and differences between groups. Comparing both the empirical and normative treatments with the baseline group, significance tests fail to reject that differences in proportions are zero for all preference types. Further, the analysis of preference types based on 'how compatible' subjects are with each type also fails to produce any significant differences. Therefore, as with H1 and H2, the conclusions of the construal level theory tests are that differences in choices between the treatments and the baseline are limited to differences in the proportions of actions chosen by subjects in the empirical and baseline groups.

6.5 Norms and Focussing

N&F1 is that there is more matching between predictions and choices in the empirical treatment than the normative treatment. Confirmation of this hypothesis would lend support to subjects reciprocating what they expect others to choose. Proportions of matching between predicted actions and chosen actions are shown in Table 6. Testing this hypothesis requires a direct comparison between the two treatment groups based on their rates of matching. This can be performed using a one-tailed Fisher exact test and the null hypothesis is that proportions of matched actions are either the same for both treatments, or subjects in the normative treatment display higher rates of matching. The alternative hypothesis is that rates of matching are significantly higher for the empirical group than the normative group. Performing this test, the null is rejected at the 1% level and, therefore, the alternative hypothesis is accepted. This finding potentially supports that, by first predicting what someone else chooses, subjects could be focussed on the norm of reciprocity.

N&F2 and N&F3 predict that subjects in the treatment groups are focussed on pro-social norms. First, N&F2 predicts that there are higher proportions of C choices in the treatment groups than the baseline group. These proportions are shown in Table 4. The null hypothesis is that the baseline group has proportions of C choices that are either the same or higher than the treatment groups. The alternative is that treatment groups choose higher proportions of C.

This can be tested using one-tailed Fisher exact tests. First, comparing the normative treatment to the baseline, the test fails to reject the null hypothesis at the 10% level. Therefore, N&F2 is rejected for the normative treatment. Second, comparing the empirical treatment to the baseline group, again the tests fail to reject the null hypothesis. Moreover, the reverse hypothesis – that proportions of C choices in the baseline are higher than the empirical treatment – is valid at the 1% level. Contrary to the prediction of N&F2, this appears to indicate that first making a prediction about someone else focusses subjects on being less pro-social.

The null hypotheses for N&F3 are that the percentages of subjects most compatible with SW and IA types are higher in the treatment groups than the baseline. These proportions are shown in Table 5. For all tests, the one-tailed Fisher exact test is used. First testing for differences in joint proportions, the null hypothesis is that the baseline has the same or higher combined proportions of SW and IA types. For both the empirical and normative treatments, the tests fail to reject the null at the 10% level. Therefore, combined proportions of SW and IA types do not significantly differ between the treatments and the baseline. Looking at SW and IA types in isolation, the null hypothesis is that the baseline has equal or higher proportions of each type than the treatments. Again, for both treatment groups, the tests fail to reject the null for both SW and IA types. As such, N&F3 is rejected for both treatment groups.

The final two norms and focussing hypotheses relate to Krupka and Weber's (2009) suggestion that the normative treatment draws subjects' attention more directly to pro-social norms the empirical treatment. As a result, N&F4 predicts the normative treatment to have a higher proportion of C choices and N&F5 predicts the normative treatment to have a higher proportion of SW and IA compatible preference types.

First investigating N&F4 using one-tailed Fisher exact tests, the null hypothesis is that the empirical group chooses an equal or greater proportion of C choices than the normative group. These proportions are shown in Table 4. The tests find the null to be rejected at the 5% level. This confirms N&F4, and therefore supports that the normative treatment provides a stronger pro-social focus than the empirical treatment.

For N&F5, the null hypothesis is that SW and IA types are either jointly or individually represented equally or better in the empirical treatment compared to the normative treatment. These proportions are shown in Table 5. Both jointly and individually, one-tailed Fisher exact tests fail to reject the null. Therefore, the only support for the normative treatment providing a more pro-social focus comes from the confirmation of N&F4.

6.6 Other Noteworthy Findings

This subsection details some additional findings not covered by the formal 'prediction then choice' hypotheses.

6.6.1 Baseline Group Choices Differ From I&R's experiment

It is interesting to compare how the results from this experiment compare to I&R's, particularly with regards to the baseline group which is intended to be a neutral comparison. Table 7 provides a summary of choices and differences between the baseline group and I&R's experiment.

Table 7 - Differences Between the Baseline and I&R's Experiment

		Proportion of choices		Difference	
		Baseline	I&R	Baseline vs. I&R	
Choice		(1)	(2)	(3)	
Action type	S	0.31	0.72	-0.41	***
	C	0.63	0.20	0.43	***
	D	0.06	0.08	-0.02	**
Preference type	SF	0.09	0.47	-0.38	***
	SW	0.68	0.19	0.49	***
	IA	0.30	0.21	0.09	
	CP	0.05	0.13	-0.08	*

Notes: Columns 1&2 display the baseline group and I&R's experiment by their compositions of S, C and D proportions, given to 2 decimal places. Column 3 shows the differences in proportions between the baseline group and I&R's experiment for each action and preference type. Differences are tested for significance using two-tailed Fisher exact tests. The null hypothesis for each test is that the difference is equal to zero. Asterisks indicate the level of significance at which the null hypothesis is rejected.

*** Significant at the 1% level
** Significant at the 5% level
* Significant at the 10% level

Differences in proportions of S, C and D choices are all significant. Similarly, all differences in preference types are significant except for the IA type. Moreover, proportions of SF and SW types appear almost reversed. Consequently, this raises questions about the methodological differences between this experiment and I&R's which appear to result in subjects choosing different arrays of social preferences. This finding is explored further in Section 7.

6.6.2 Baseline Group Predictions Differ from Choices

Section 6.6.1 establishes that choices in the baseline group do not replicate I&R's findings. As such, it is worth examining the premise of the baseline treatment that the predictions are trivial

and should perfectly match choices. Table 8 reports the differences in baseline predictions and choices. There are no significant differences in preferences types, however, the proportions of S and C actions differ at the 5% and 1% levels respectively. Indeed, one-tailed Fisher exact tests confirm that subjects in the baseline group predicts that, one hour ago, they would have played a higher frequency of S choices and a lower frequency of C choices than in the present moment. This indicates that people behave more altruistically than they predict they would have behaved in the 'near' past. Unexpectedly, this contradicts the premise of the baseline treatment that making a prediction of own choices one hour ago would be trivial. Consequently, it follows that the baseline group is not perfectly suited for providing a neutral basis for comparison with the treatments. However, in the context of this experiment, it remains the most appropriate basis for comparison.

Table 8 - Differences Between Baseline Predictions and Choices

		Baseline		Difference	
		Prediction	Choice	Prediction vs. Choice	
Choice		(1)	(2)	(3)	
Action type	S	0.38	0.31	0.06	**
	C	0.56	0.63	-0.07	***
	D	0.06	0.06	0.01	
Preference type	SF	0.14	0.09	0.05	
	SW	0.59	0.68	-0.09	
	IA	0.31	0.30	0.01	
	CP	0.11	0.05	0.06	

Notes: Columns 1&2 display the baseline group's predictions and choices by their compositions of S, C and D proportions, given to 2 decimal places. Column 3 shows the differences in proportions between predictions and choices for each action and preference type. Differences are tested for significance using two-tailed Fisher exact tests. The null hypothesis for each test is that the difference is equal to zero. Asterisks indicate the level of significance at which the null hypothesis is rejected.

*** Significant at the 1% level

** Significant at the 5% level

* Significant at the 10% level

6.6.3 Demographic Indicators

Finally, it is valuable to investigate the relationships between choices and demographic variables. Table 9 displays the results of regressions with action types as the dependent variables. Action types are continuous variables with values ranging from 0 to 8 based on how many times they were chosen. Most independent variables do not report significant correlations with action types. However, a noteworthy exception is the binary variable 'EMAIL' which indicates

whether a subject choses to provide their email address at the end of the experiment in order to be entered into the lottery for payment. Ceteris paribus, leaving an email address is associated with playing: 1.1 more C choices (significant at the 1% level), and 0.38 fewer D choices (significant at the 5% level). This finding is interesting because it might be expected that those playing the highest frequencies of S choices, rather than C, stand to gain most by leaving their email address. A potential explanation is that those willing to share more with others are also more likely to share their personal information with the experiment.

Table 9 - Relationships Between Action types and Demographic Indicators

	S		C		D	
	(1)		(2)		(3)	
Empirical	0.49		-0.49		0.01	
	(0.35)		(0.36)		(0.16)	
Normative	0.25		-0.13		-0.03	
	(0.35)		(0.38)		(0.15)	
AGE	0.00		0.00		0.00	
	(0.01)		(0.01)		(0.01)	
MALE	0.36		0.13		-0.13	
	(0.30)		(0.32)		(0.13)	
STDNT	0.48		-0.75	*	0.13	
	(0.37)		(0.39)		(0.17)	
ECON	0.55	*	-0.48		-0.06	
	(0.32)		(0.34)		(0.14)	
EMAIL	-0.56		1.10	***	-0.38	**
	(0.39)		(0.36)		(0.21)	
Constant	2.21	***	4.58	***	0.70	**
	(0.70)		(0.64)		(0.30)	
Observations	267		267		267	
R^2	0.047		0.074		0.032	

Notes: Columns 1-3 display the results from OLS regressions with action types as dependent variables. The action types are continuous variables ranging in value from 0 to 8 based on how many times they were chosen. The independent variables are all categorical except for AGE which is continuous. Empirical = 1 if a subject is in the empirical treatment. Normative = 1 if a subject is in the normative treatment. MALE = 1 if a subject is male. STDNT = 1 if a subject's primary occupation is studying. ECON = 1 if a student has completed a university level economics course. EMAIL = 1 if a subject elects to provide their email address. The variables JOB (a subject is full or part-time employed), HEDUC (a subject has completed some higher education), and TIME (continuous measure of time taken to complete the survey) were included in an earlier specification of the model. However, there is no theoretical basis to include these variables, and they were omitted after tests for joint significance rejected that they had any significant explanatory power. Coefficients and robust standard errors are given to two decimal places. Significance is indicated by asterisks.

*** Significant at the 1% level
** Significant at the 5% level
* Significant at the 10% level

7 Discussion

This section presents a discussion of the results' main findings.

Finding 1: Anchoring and adjustment cannot be ruled out.

Weak evidence of anchoring and adjustment is indicated by rates of matching between predictions and choices that are stronger than chance. This finding confirms AA1 and indicates that, at the very least, a process of anchoring and adjustment cannot be ruled out. However, the additional condition proposed by AA2 – that rates of matching between predicted actions and chosen actions are equal across groups – is rejected. Therefore, although anchoring and adjustment cannot be ruled out, it is apparent that its influence, if any, varies based on the type of prior prediction a subject makes.

The support for anchoring and adjustment in this experiment, albeit weak, corroborates the support found for anchoring and adjustment in the context of risk preferences (Li, Rohde & Wakker, 2017). However, the tests for anchoring and adjustment in this experiment are not ideal. The support for AA1 only allows for the conclusion that anchoring and adjustment cannot be ruled out, whereas AA2 is strict by assuming that anchoring and adjustment produces a dominant influence in choice decisions that is equal across groups. In future experiments, a more appropriate method for identifying the presence of anchoring and adjustment could be to include additional groups who first make all predictions and then make all choices, perhaps with a task or time in between. Lower rates of matching for these groups compared to the corresponding original 'prediction the choice' groups would be a better indicator of whether the predictions made immediately prior to choosing are anchored upon.

Finding 2: There is no evidence of a false consensus effect.

Evidence of a false consensus effect rests on two premises. First, like AA1, rates of matching between predictions and choices are stronger than chance. Second, there are no significant differences in choices between groups. Together, these suggest that subjects expect others to behave as they do, with the treatments having no effect on choices. In particular, the second premise is central to the false consensus hypothesis. This is because the theory proposes that predictions about others derive from a subject's own position, therefore the type of prior prediction should have no effect on choices.

The weak support for anchoring and adjustment means that FC is not rejected based on the first premise. However, the results show that FC is rejected based on the second premise. There are no significant differences in the proportions of action types chosen between the normative treatment and the baseline group; however, the empirical treatment differs from both groups in terms of the proportions of S and C choices. As such, it is apparent that the type of prior prediction a subject makes does have an influence on the choices they make for themselves. Thus, the contradiction of the second premise means FC is rejected.

Contrary to the present findings, I&R propose that their results indicate the presence of a false consensus effect. They base this finding on the premise that a false consensus effect is shown by, for example, selfish types predicting others to act selfishly. Whereas, in this experiment, the use of different treatments means it is important to factor in the second premise that choices are unaffected by the type of prior prediction a subject makes. Based on this stricter test of the false consensus effect, it is valid to reject FC under this experimental framework.

Finding 3: Compared to the baseline, subjects in the empirical treatment choose S more frequently and C less frequently.

The tests of the general hypotheses verify that the empirical treatment does produce choices which differ from the other groups. H1 – actions differ between groups – is confirmed for the empirical treatment given that proportions of S and C choices differ from both the baseline and normative groups. Although, H2 – preference types differ between groups – is not supported for the empirical treatment using either the 'most compatible' or 'how compatible' classifications of preferences types. However, the support for H1 indicates that, in general, subjects behave more selfishly when they are first asked to predict the actions of someone else.

To help explain these findings, Table 3 shows that the difference between the normative group's predicted and chosen proportions of S and C are 14 and 16 percentage points respectively, practically double the 6 and 7 percentage points difference for the same predicted and chosen actions in the empirical treatment. This could indicate that anchoring has a stronger influence in the empirical treatment than the normative treatment. Since all groups predict higher proportions of S choices than they choose – and S predictions in the empirical treatment are higher than the baseline – a stronger influence of anchoring helps to explain why the empirical group produces more selfish choices than both the normative group and the baseline. Although, Table 2 shows that subjects in the empirical group predict S and C at comparable rates, yet there is roughly a 16-percentage points difference in the rates at which these actions

are chosen. This appears to be a strong indication that anchoring cannot be the only influence that predictions have on choices.

Therefore, other behavioural theories are important to factor in. Construal level theory proposes that increasing psychological distance will change the nature of a subject's decision making. Between the treatments and the baseline, CLT1 is that choices will differ and CLT2 is that preference types will differ. Similar to H1, CLT1 is confirmed for the empirical group through proportions of S and C choices differing significantly from the baseline. However, CLT2 is rejected when comparing between the empirical treatment and the baseline for both 'most compatible' and 'how compatible' classifications of preference types. In addition, N&F2 predicts that, compared to the baseline, subjects in the treatment groups make more pro-social (C) choices. Not only is N&F 2 rejected for the empirical treatment, the reverse hypothesis – that subjects in the empirical treatment make fewer pro-social choices – is supported at the 1% level.

Consequently, it is apparent that first making a prediction of what someone else chooses results in subjects making choices that are more selfish and less altruistic than they would be otherwise. Considering construal level theory, this would mean the effects of psychological distance result in more self-interest; which would be consistent with the notion that altruistic decisions are influenced by emotions. In terms of the norms and focussing theory, rather than the empirical treatment focussing subjects on pro-social norms, perhaps subjects are instead focussed on a norm of self-interest.

This finding strongly contradicts those of Krupka and Weber (2009) who observe subjects to act more pro-socially after first predicting what someone else would do. However, methodological differences between the two experiments could be a factor. One major difference is that subjects in Krupka and Weber's experiment made one choice over two outcomes, whereas subjects in this experiment made ten choices over three outcomes. In particular, the addition of a spiteful (D) choice may have influenced how subjects approach the S and C choices. It is possible that subjects perceive the D choice as 'asymmetrically dominated' by the S choice given that neither option is pro-social, but the S choice is less anti-social. According to research by Huber, Payne and Puto (1982), adding an asymmetrically dominated option to a choice list can increase the attractiveness of the option that dominates it. Therefore, it is possible that the higher proportions of selfish choices could partly be the product of the choice architecture.

Further, the additional decisions required for this experiment induces a greater demand on cognitive resources, which could bear some responsibility for the observed differences in preferences. Another factor worth considering is that subjects made 7 out of 8 predictions directly after making a choice, albeit for a different table. As such, it is possible that some residual influence from previous choices spilled over onto the decisions made for subsequent predictions. Unfortunately, subjects made decisions on the tables in a random order, so it is not possible to see if the decisions made for the first table differ from other decisions.

Finding 4: Compared to the baseline, subjects in the normative treatment do not choose significantly different actions.

Unlike the empirical treatment, both H1 and CLT1 are rejected for the normative treatment since, compared to the baseline group, there are no significant differences between proportions of chosen actions. Further, based on both 'most compatible' and 'how compatible' classifications of preference types, the results fail to support H2 and CLT2. In addition, N&F2 is also rejected since subjects in the normative treatment do not make more pro-social choices than the baseline group. Therefore, after first predicting what others would say one should do, there is no strong evidence to suggest that choices differ from the baseline group. This result also contradicts the findings of Krupka and Weber (2009) because the lack of difference defies the expectation that subjects in the normative treatment choose more altruistic actions and behave more in keeping with pro-social preference types. Further, unlike the empirical treatment, there is no evidence to support that there is an effect of psychological distance when first making a prediction about what others would say one should do. A possible explanation for this is that the predictions made by subjects in the normative treatment are reflexive[11], and so do not cause the subjects to experience psychological distance through focussing on others. Overall, the finding that the normative treatment does not produce different chosen actions or preference types to the baseline group reflects Bicchieri and Xiao's (2009) findings that normative beliefs have no significant effects on personal choices.

Comparing predictions to choices at the group level (Table 3), it is also interesting to note that the normative prediction produces the highest differences for any group between predicted and chosen S and C actions and SF and SW preference types. This is also in keeping with the normative treatment producing the lowest rate of matching between predictions and choices,

[11] The normative treatment asks subject to predict what others would say *the subject* should choose. Hence, the ultimate focus remains centred on the self.

with the average subject matching 64% compared to rates of 85% and 75% for the baseline and empirical groups respectively (Table 6). Consequently, compared to the empirical treatment, the weaker influence of anchoring and adjustment is likely to be one explanation for selfish predictions not being replicated in choices.

Finding 5: Subjects in the empirical treatment could be focussed on reciprocity.

Another hypothesis derived from the norms and focussing theory is that subjects in the empirical treatment are more focussed on the norm of reciprocity than subjects in the normative treatment. Concurrently, N&F1 is that, compared to the normative treatment, there are higher rates of matching between predictions and choices in the empirical treatment which significance tests confirm to be valid. This finding seems to support subjects reciprocating what they expect others to choose. Therefore, in the empirical treatment, if subjects' perceptions of what others choose were to be influenced, this could translate to subjects making different choices. Potentially, this would be an interesting area for future study.

Finding 6: Compared to the empirical treatment, subjects in the normative treatment choose a higher proportion of C choices.

Krupka and Weber (2009) suggest that the normative treatment more directly focusses subjects on pro-social norms than the empirical treatment. Correspondingly, N&F4 predicts the normative treatment to have a higher proportion of C choices and N&F5 predicts the normative treatment to have a higher proportion of SW and IA compatible preference types. The results show N&F4 to be valid at the 5% level, whereas N&F5 is rejected.

Initially, this finding seems to concur with Krupka and Weber's suggestion. However, given that the baseline group also produces more pro-social choices than the empirical group, it could also be argued that, rather than focussing subjects on pro-social norms, the normative treatment simply provides less of a self-interested focus than the empirical treatment. Indeed, since the chosen actions and preference types are shown not to differ significantly from the baseline, it is apparent that neither treatment influences subjects to be pro-social. Consequently, these results recommend that organisations seeking charitable donations do not ask people to first predict what others do or believe. On the other hand, if another situation requires people to act more selfishly, they should first be asked to predict the action someone else would take.

Finding 7: Choices made by the baseline group differ from I&R's experiment.

Comparing the choices made by the baseline group to I&R's experiment reveals significant differences in the proportions of all action types and 'most compatible' preference types except for IA. Unexpectedly, this highlights that the different experimental methodologies used have resulted in subjects making distinctly different choices. Contributing factors in this experiment could be, for example, the minimal real incentives offered, the online survey format, and differences in subject demographics – although, most demographic indicators were not found to be significant by the regression analysis (Table 8).[12]

However, one key difference between the two experimental methodologies is that I&R's experiment uses role certainty – that is, Deciders know for certain that they are making choices for themselves and their partnered Receiver. Whereas, in this experiment, all subjects make choices as a Decider. It transpires that, in an earlier paper, Iriberri and Rey-Biel (2011) use the same experimental framework to compare both role certainty and role uncertainty for Deciders[13]. Their role uncertainty group chooses S, C and D at rates of 0.35, 0.64 and 0.02 respectively – close to matching the 0.31, 0.63 and 0.06 baseline choice rates in this experiment. However, their role certainty group chooses S, C and D at rates of 0.69, 0.24 and 0.08 which are roughly the proportions replicated in their later study (I&R Part 1). Despite how, in theory, subjects should make identical choices regardless of whether their role is certain, Iriberri and Rey-Biel's (2011) findings show how role uncertainty produces distinctly different choices. Further, these differences persist even when tests are used to check for subjects' understandings of the payoff mechanisms. As a result, Iriberri and Rey-Biel (2011) conclude that role uncertainty seems to increase altruism and reduce the presence of selfish and spiteful behaviour.

Given these findings, using role uncertainty appears to explain most of the variation between the experimental findings in this paper and I&R Part 1. For future study, replications of this experiment using role certainty would be interesting. Despite this, using role uncertainty in the present experiment does not invalidate findings for two reasons. First, the main intention of this experiment was to compare the effects of treatments to each other. Since role uncertainty was used for all treatment groups, comparisons are neutral and are, therefore, still valid. Second, there are some instances of role uncertainty in everyday life that these findings may be more applicable to. For example, voters cannot know whether their vote is for the winning party or not when it is cast. Assuming they still vote, it is possible that voters would make

[12] Differences in the mapping of subjects to behavioural rules is unlikely to be a major factor since there are also pronounced differences in the chosen actions.
[13] I discovered this paper after the experiment had already been completed.

different choices if they knew for sure what the outcome of an election would be. Additionally, had role certainty been used, only half the data would have been gathered, considerably reducing the power of the experimental tests.

Finding 8: Baseline group predictions differ from choices.

The premise of the baseline group is that, in most cases, subjects would choose the same actions one hour ago as they do in the present moment. Indeed, rates of matching between predictions and choices was 85%, notably higher than the 75% in the empirical treatment and the 64% in the normative treatment. However, it appears the baseline group predicts that, one hour ago, they would have played a higher frequency of S choices and a lower frequency of C choices than in the present moment. This implies that people view themselves as less altruistic in the immediate short-term. Given this finding, in future study, it might prove interesting to use treatments where subjects predict how they would have made their decisions with varying degrees of retrospect.

However, the differences between predictions and choices raises concerns over the suitability of the baseline group as a basis for neutral comparison with the treatments. Although, the challenge lies in proposing a more suitable alternative. If the baseline had asked only for choices, this would mean that subjects answer half as many questions as those in the treatment groups and a separate set of instructions would be required. Further, tests for AA2 would have been limited to comparisons between treatments. As such, the current baseline may prove more suitable than any other alternative.

8 Conclusion

This study seeks to bridge a gap in the research concerning how the effect of 'prediction then choice' influences social preference decisions and the subsequent categorisation of subjects into four distinct social preference types. Analysing the results reveals several findings. First, there are significant differences in the types of actions chosen by subjects in the empirical treatment compared to the other groups. Therefore, prior predictions do influence social preferences. Further, since choices are influenced by the types of predictions made about others, this rules out a false consensus effect.

Second, there is evidence that rates of matching between predictions and choices are stronger than chance. As such, a process of anchoring and adjustment cannot be ruled out. Although, since rates of matching differ between groups, the influence of anchoring and adjustment is likely to depend on the type of prior prediction that is made.

Third, subjects in the empirical treatment make more selfish choices and fewer surplus creating choices than both the normative treatment and the baseline. This finding suggests that people are more selfish and less altruistic when first making a prediction about someone else's actions. As a result, this contrasts with Krupka and Weber's (2009) findings, and the predictions of the norms and focussing theory, that subjects in the empirical treatment would make more pro-social choices than the baseline. On the other hand, this confirms the prediction of construal level theory that increasing 'psychological distance' results in different choices.

Fourth, comparing the normative treatment with the baseline, there are no indications of significant differences between choices or preference types. Consequently, these findings also disagree with Krupka and Weber's (2009) findings and the norms and focussing theory. In addition, the findings for the normative treatment offer no support for construal level theory. As such, this experiment finds no evidence that 'normative' prior predictions influence social preferences.

The remaining findings open possibilities for future study. First, there is some evidence that the norm of reciprocity is strongest for subjects in the empirical treatment. As such, it is plausible that people's social preferences could be influenced by changing their perceptions of how they expect others to act. Second, since there are considerable differences between the actions subjects choose and the actions they predict themselves to have chosen one hour ago, it may prove worthwhile to test varying degrees of short-term retrospection as treatments in their own

right. Finally, the conclusions of Iriberri and Rey-Biel (2011) suggest that the role uncertainty used in this experiment may have resulted in subjects making substantially more pro-social choices. For that reason, there is a case for replicating this study in the future using role certainty; although, it is reasonable to assume that comparisons between treatments remain valid for the present study.

Overall, no significant differences are found between the chosen preference types of any groups. This is despite using both 'most compatible' and 'how compatible' measures of preference type compatibility. Nonetheless, the result that neither treatment group displays more pro-social preference types than the baseline group is incompatible with the findings of Krupka and Weber (2009). Moreover, it is remarkable that the empirical treatment not only fails to produce pro-social behaviour, but significantly increases proportions of selfish choices. In summary, it is surprising that the results in this paper strongly contradict the findings of Krupka and Weber (2009), especially since the intention of this study was to advance their experiment using a different method for eliciting social preference types. Accordingly, it seems this contradiction stands as a further testament to the complex nature of social preferences.

In terms of the practical relevance of these findings, there are situations where both selfishness and altruism could be encouraged. For example, encouraging those who overburden themselves at work to behave more selfishly could reduce risks of burnout. On the other hand, altruistic behaviour helps less fortunate others, and could provide mutual benefits in situations where reciprocity is expected. Overall, when selfishness is desired, the findings recommend that people first predict the actions of someone else before making their own choice. However, when altruism is desired, the findings recommend that it is best to avoid subjects predicting what others do, or say one should do.

References

Andreoni, J. & Miller, J. (2002). Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica*, *70*, 737–753.

Bernheim, B. D. (1994). A theory of conformity. *Journal of Political Economy*, *102*(5), 841-877.

Bicchieri, C., & Xiao, E. (2009). Do the right thing: but only if others do so. *Journal of Behavioural Decision Making*, *22*, 191-208.

Blanco, M., Engelmann, D., & Normann, H. (2011). A within-subject analysis of other regarding preferences. *Games and Economic Behavior*, *72*(2), 321–338.

Blanco, M., Engelmann, D., Koch, A. K., & Normann, H. T. (2014). Preferences and beliefs in a sequential social dilemma: a within-subjects analysis. *Games and Economic Behavior*, *87*, 122-135.

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

Cabrales, A., Miniaci, R., Piovesan, M., & Ponti, G. (2010). Social preferences and strategic uncertainty: an experiment on markets and contracts. *American Economic Review*, *100*(5), 2261-2278.

Chapman, G. B., & Johnson, E. J. (1999). Anchoring, activation, and the construction of values. *Organizational behavior and human decision processes*, *79*(2), 115-153.

Chapman, G. B., & Johnson, E. J. (2002). Incorporating the irrelevant: Anchors in judgments of belief and value. *Heuristics and biases: The psychology of intuitive judgment*, 120-138.

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

Chaxel, A. S. (2014). The impact of procedural priming of selective accessibility on self-generated and experimenter-provided anchors. *Journal of Experimental Social Psychology*, *50*, 45-51.

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

Costa-Gomes, M. & Weizsäcker, G. (2008). Stated beliefs and play in normal form games. *Review of Economic Studies*, *75*, 729-762.

Croson, R. T. A. (2000). Thinking like a game theorist: factors affecting the frequency of equilibrium play. *Journal of Economic Behaviour & Organization*, *41*, 299-314.

Dawes, R. M. (1989). Statistical criteria for establishing a truly false consensus effect. *Journal of Experimental Social Psychology*, *25*, 1 –17.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, *14*(4), 583-610.

Engelmann, D., & Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review*, *94*(4), 857-869.

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science*, *17*(4), 311-318.

Faro, D., & Rottenstreich, Y. (2006). Affect, empathy, and regressive mispredictions of others' preferences under risk. *Management Science*, *52*, 529–541.

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

Fisman, R., Kariv, S., & Markovits, D. (2007). Individual preferences for giving. *American Economic Review*, *97*(2), 153–158.

Gächter, S., & Renner, E. (2010). The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics*, *13*(3), 364-377.

Galizzi, M. M., & Navarro-Martínez, D. (2018). On the external validity of social preference games: a systematic lab-field study. *Management Science*, *65*(3), 976-1002.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive psychology*, *24*(1), 1-55.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of consumer research*, *9*(1), 90-98.

Iriberri, N., & Rey-Biel, P. (2011). The role of role uncertainty in modified dictator games. *Experimental Economics*, *14*(2), 160-180.

Iriberri, N., & Rey-Biel, P. (2013). Elicited beliefs and social information in modified dictator games: What do dictators believe other dictators do? *Quantitative Economics*, *4*, 515-547.

Johnson, E. J., Häubl, G., & Keinan, A. (2007). Aspects of endowment: a query theory of value construction. *Journal of experimental psychology: Learning, memory, and cognition*, *33*(3), 461.

Johnson, E. J., Shu, S. B., Dellaert, B. G., Fox, C., Goldstein, D. G., Häubl, G., ... & Wansink, B. (2012). Beyond nudges: Tools of a choice architecture. *Marketing Letters*, *23*(2), 487-504.

Kerschbamer, R., Sutter, M., & Dulleck, U. (2016). How social preferences shape incentives in (experimental) markets for credence goods. *The Economic Journal*, *127*(600), 393-416.

Krupka, E., & Weber, R. A. (2009). The focusing and informational effects of norms on pro-social behavior. *Journal of Economic Psychology*, *30*(3), 307-320.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of economic dynamics*, *1*(3), 593-622.

Li, Z., Rohde, K. I. M., Wakker, P. P. (2017). Improving one's choices by putting oneself in other's shoes – An experimental analysis. *Journal of Risk and Uncertainty*, *54*(1), 1-13.

Liberman, N., Trope, Y., & Wakslak, C. (2007). Construal level theory and consumer behavior. *Journal of Consumer Psychology*, *17*, 113 –117.

Marks, G., & Miller, N. (1987). Ten years of research on the false-consensus effect: An empirical and theoretical review. *Psychological bulletin*, *102*(1), 72.

Moore, D. A. (1999). Order effects in preference judgments: Evidence for context dependence in the generation of preferences. *Organizational behavior and human decision processes*, *78*(2), 146-165.

Paternoster, R., Jaynes, C. M., & Wilson, T. (2017). Rational choice theory and interest in the "Fortune of Others". *Journal of research in crime and delinquency*, *54*(6), 847-868.

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect". An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 13, 279-301.

Rutström, E. E., & Wilcox, N. T. (2009). Stated beliefs versus inferred beliefs: A methodological inquiry and experimental test. *Games and Economic Behavior*, *67*(2), 616-632.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, 117(2), 440.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1130.

Appendix

1. Experiment instructions

Thanks for taking part in this experiment.

In total, the estimated completion time is 7-9 minutes but please take as long as you need.

Some of the choices you make have the chance to be played out for real. This means your decisions could earn you money.

If you do not want the chance of being paid, please complete the experiment anyway but do not include your email address at the end of the experiment.

Your answers and any personal information will be used for this experiment only.

Payments

Each participant has an equal chance of their choices being paid out for real.

This experiment uses euros, but you can choose to be paid in any currency you prefer.

Instructions

Please read the instructions carefully.

In this experiment, there are two roles: Decider and Receiver. The Decider chooses pay-offs for both roles. You will be shown tables like this:

	Option 1	Option 2	Option 3
Decider	€8	€7	€11
Receiver	€17	€19	€13

If the Decider chooses Option 1, (s)he would get €8 and the Receiver would get €17.

If the Decider chooses Option 2, (s)he would get €7 and the Receiver would get €19.

If the Decider chooses Option 3, (s)he would get €11 and the Receiver would get €13.

For each table you are presented with, you will be anonymously paired with another participant. Neither of you will ever find out the other's identity or their choices.

When you make your decisions, you are always the Decider and your partner is the Receiver. So by choosing Option 1, you would get €8 and your partner would get €17.

When your partner makes his/her decisions, (s)he takes the role of Decider and you are the Receiver. For example, if your partner chooses Option 1, (s)he would get €8 and you would get €17.

Payments

In total, there are 10 tables each with a Part A and Part B question. Only Part B questions have the chance to be played for real.

Across the whole experiment, one partnership will have one of the Part B questions played for real. If your pair is chosen, a coin is flipped to determine whether your Decider choices or your partner's Decider choices determine the final payments.

The experiment begins on the next page. There are 10 tables in total.

This experiment is simply about understanding your preferences over payoffs for yourself and someone else. Pairing you with another participant is only necessary to ensure that your choices can be paid out for real if you are selected. There are no right or wrong answers.

---

2. Decision Table 1: Baseline

Please consider the following table

	Option 1	Option 2	Option 3
Decider	€7	€7	€8
Receiver	€10	€24	€17

Part A

On the next page, Part B will ask you to make your choice as the Decider.

First, please select the option you think you would have chosen 1 hour ago.

- Option 1
- Option 2
- Option 3

You are the Decider and your partner is the Receiver

	Option 1	Option 2	Option 3
Decider	€7	€7	€8

	Option 1	Option 2	Option 3
Receiver	€10	€24	€17

Part B

Please select your preferred payoff option. Your choice could be played out for real.

- Option 1
- Option 2
- Option 3

---

3. Decision Table 1: Empirical Treatment

Please consider the following table

	Option 1	Option 2	Option 3
Decider	€7	€7	€8
Receiver	€10	€24	€17

Part A

On the next page, Part B will ask you to make your choice as the Decider.

First, consider that your partner also faces an identical decision as a Decider in Part B. Please predict the option you think your partner is most likely to choose as a Decider.

- Option 1
- Option 2
- Option 3

You are the Decider and your partner is the Receiver

	Option 1	Option 2	Option 3
Decider	€7	€7	€8
Receiver	€10	€24	€17

Part B

Please select your preferred payoff option. Your choice could be played out for real.

- Option 1
- Option 2
- Option 3

---

4. Decision Table 1: Normative Treatment

Please consider the following table

	Option 1	Option 2	Option 3
Decider	€7	€7	€8
Receiver	€10	€24	€17

Part A

On the next page, Part B will ask you to make your choice as the Decider.

First, please select the option that you believe others would say you should choose.

- Option 1
- Option 2
- Option 3

You are the Decider and your partner is the Receiver

	Option 1	Option 2	Option 3
Decider	€7	€7	€8
Receiver	€10	€24	€17

Part B

Please select your preferred payoff option. Your choice could be played out for real.

- Option 1
- Option 2
- Option 3