MICHIGAN
ROSS SCHOOL OF BUSINESS

Working Paper

Systematic Differences in Beliefs about Others in Strategic Interactions

Yesim Orhun
Stephen M. Ross School of Business
University of Michigan

Ross School of Business Working Paper
Working Paper No. 1185
July 2010

This work cannot be used without the author's permission.
This paper can be downloaded without charge from the
Social Sciences Research Network Electronic Paper Collection:
http://ssrn.com/abstract=2241702

UNIVERSITY OF MICHIGAN

Systematic Differences in Beliefs about Others in Strategic Interactions*

A. Yeşim Orhun†

This version: July 2010.
First version: July 2009.

Working paper version, please do not distribute or cite without permission.

Abstract

Strategic interactions are governed both by individuals' preferences for outcomes and by their expectations about other players' choices that influence the outcome . The common assumption that expectations about others are mutually consistent across players allows researchers to infer preferences from observed strategic decisions. In this paper, I show how players' beliefs about other players' choices systematically depart from this assumption and I explain the consequences for inferring preferences based on strategic choices.

In the context of altruistic preferences, I document a relationship between an individual's preferences and his (implicit or explicit) expectations of others' actions in modified dictator games. This relationship is beyond what false consensus or a simple correlation between beliefs and preferences can account for and is consistent with a more complex relationship of beliefs and preferences that underlie choices.

I study the impact of these systematic belief differences on players' strategic actions in a trust-dictator game. I show that preference incongruencies across different roles in a trust-dictator game are in line with the documented relationship between beliefs and preferences. Finally, I demonstrate biases in the estimation of preferences from decisions in this strategic game under the assumption of mutually consistent beliefs.

1 Introduction

Utility-maximizing decision makers consider two factors when choosing the optimal action in any strategic interaction: their evaluation of potential outcomes and the likelihood of reaching each outcome given that action. The perceived likelihood of reaching an outcome depends on the decision maker's expectations about others' actions. As researchers generally do not observe these beliefs, empirical inference of preferences from observed strategic choices predominantly relies on the as-

*I am indebted to Oleg Urminsky for his invaluable support and feedback. I thank Allen Bester, Stefano DellaVigna, Uri Gneezy, Christian Hansen, Reid Hastie, Shachar Kariv, George Wu, Florian Zettelmeyer, Ting Zhu, and the participants at seminars in Columbia University, Washington University at St. Louis, University of Chicago, UC San Diego and in the Marketing in Israel Conference for their helpful comments. Financial support from the James M. Kilts Center for Marketing at the University of Chicago Booth School of Business is gratefully acknowledged.

†The University of Chicago, The Booth School of Business. 5807 S Woodlawn Avenue, Chicago, IL 60637. yesim.orhun@chicagobooth.edu.

sumption that beliefs about others' actions are objectively correct (i.e., rational[1]). Manski (2002, 2004) points out the identification concern associated with relying on this potentially misspecified assumption in order to separate preferences from expectations in choice data, and he advocates measuring expectations.

The assumption that expectations are objectively correct includes two premises: first, beliefs on average reflect the realized outcomes; and second, no systematic variation in expectations exists. Whereas the first premise provides separation of latent average preferences from latent average beliefs, the second premise (mutual consistency assumption) allows any variation in choices across people or time to be attributed to variation in preferences. Recent literature in economics has considered the fact that mean beliefs may systematically deviate from reality (i.e., overconfidence (Camerer and Lovallo (1999), Barber and Odean (2001), Malmendier and Tate (2008)) and conservatism (Huck and Weizscker (2002)). This research on inaccuracy of beliefs mostly focuses on the violations to the first premise and does not directly speak to the possibility that different individuals may have different beliefs, regardless of how mean beliefs compare with reality. Individual random deviations from common beliefs may not hamper the identification of a ranking in preferences. However, if beliefs are systematically different across people in a way that relates to their preferences that influence their actions then we can no longer preference-rank individuals based on the differences in their actions. Previous studies in psychology and economics document that a person who has taken a particular action believes that action to be more prevalent among others than a person who has taken an alternative action (i.e., false consensus (Ross et al. (1977), Marks and Miller (1987); for review, see Selten and Ockenfels (1998), Charness and Grosskopf (2001)), which gives us reason to doubt the assumption of mutual consistency.

This paper documents the existence of systematic belief differences across people, discusses how these differences relate to differences in preferences, and studies the impact of this systematic heterogeneity on both the actions people take in strategic interactions and the inference of preferences based on such actions. The observation that the knowledge of the correct form of this relationship is essential in informing the data and statistical modeling necessary to achieve unbiased estimates of the preference distribution, even when expectations are measured, motivates the focus on the

[1] Although the less restrictive form of rational expectations assumption only requires players to hold accurate beliefs given their private information, since the heterogeneity in private information is unobserved, this more restrictive version is often used for inference.

specifics of the relationship between preferences and beliefs..

The particular application focuses on beliefs and preferences for altruism, an issue that has received significant attention in the recent experimental literature. Previous studies document considerable heterogeneity in the concern for others (for a review, see Camerer, 1993), making altruism an appealing context in which to investigate the extent of decision makers' common beliefs about this heterogeneity. Moreover, the fact that some of the behavior in social interactions that we think of as altruistic may in fact be strategic and selfish given their beliefs, noted previously by Manski (2002) and Charness and Rabin (2002) among others, further motivates the choice of this context. By focusing on individuals' degree of altruism and their beliefs about others' altruistic actions, I speak directly to these concerns.

In three experiments, I infer subjects' degree of altruism from their actions in modified dictator games. I use the linear asymmetric social preferences model introduced by Charness and Rabin (2002), distinguishing the degree of concern for others when the person has a higher payoff from when he has a lower payoff than other player. Study 1 directly elicits subjects' beliefs about others' choices in modified dictator games, and Study 2 indirectly elicits beliefs by giving subjects risky decision tasks involving gambles that pay what a random other subject from a past experiment allocated in a dictator game. In Study 3, subjects face a trust-dictator game and choose between a social-welfare minimizing allocation and deferring the decision to another person who then makes a different social allocation decision as a dictator. In this game, the decision of the first player depends both on his altruism and on his beliefs about what the second person will choose as a dictator. Therefore, the actions of the first player in this game can be interpreted as altruistic or simply strategic, depending on his beliefs.

Impact of preferences on beliefs

The first set of results document the impact of an individual's preferences on his beliefs about others' actions. I find three main patterns.

First, I find a considerable amount of heterogeneity across people in their beliefs about how others will choose in a dictator game. As expected, a person who prefers a particular outcome is more likely to believe decision makers will pick that outcome than is another person who does not prefer that particular outcome. This finding is consistent with the previous literature on false consensus, and I demonstrate its robustness to several important control factors. I show the systematic belief

differences are robust to monetary incentives for accuracy, do not depend on whether beliefs and preferences are elicited first, and are not driven by people with more inconsistencies. I show that the act of having made a choice has the same impact on belief as one's future choice or predicted choice. I also provide the first evidence that implicitly measured beliefs (inferred from decisions involving gambles on altruism) reflect systematic differences just as explicitly elicited beliefs. Finally, I also provide the first evidence of heterogeneity in beliefs in a strategic choice setting, demonstrated by incongruencies between strategic and non-strategic choices in trust-dictator games.

Second, I find that, although consistent with the previous literature on false consensus, a mere correlation of choices with beliefs falls short of capturing all the belief heterogeneity. Between two people who both favor the same option, the one who has a stronger preference for it also has a higher expectation that others will choose the same option. This novel finding is consistent with the idea that people think others have underlying preferences similar to their own. Moreover, I find that people not only assume similarity in the degree of concern for others overall but also assume similarity in the degree of concern withdrawal that results from having less rather than having more payoff than the other person. I also show that people demonstrate sensitivity to the differences in the context between themselves and others when forming beliefs about others' actions. In an interaction between Person A and B, where Person A has a higher payoff than Person B, Person A's beliefs about the choice of Person B is influenced by Person A's degree of altruism when he has the lower payoff, and not by his degree of altruism in his current position.

Third, I show that deviations from mean beliefs are note solely person-specific. If each subject stated her beliefs about others' choices (or acted on these beliefs) in only one decision task, as often is the case in the experimental studies of false consensus, discerning whether people have individual-level relative pessimism or optimism or whether their relative pessimism or optimism depends on their valuations of the choice options would be impossible. For example, those who have a higher concern for others might also be more optimistic people, resulting in a spurious person-specific relationship between beliefs and preferences. Using within-person variation in beliefs, I am able to present novel evidence that deviations from mean beliefs are not person-specific. As a person's utility difference between the choice options varies across decision tasks, so does his relative pessimism or optimism compared with the population.

Impact of belief heterogeneity on strategic actions and inference

The second set of findings expose the pattern that the relationship between preferences and beliefs creates in strategic actions, and its consequences for preference identification from these actions.

I show that if the assumption of mutual consistency is maintained, people's strategic actions in trust-dictator games imply preferences that are inconsistent with the preferences that their actions in dictator games reveal. In particular, differences between individuals regarding their trust of others are opposite of those we would expect if we assume individuals are best responding to homogenous expectations. I show that these results are congruent with differences in expectations of altruism resulting from putting one's self in the other's shoes, so to speak. In the same application, I provide evidence of the degree and direction of bias in the preference distribution estimates.

Knowing the nature of the relationship between preferences and expectations about others in strategic interactions is important in determining the necessary data and proper econometric model for the identification of preferences from a particular dataset of strategic actions. Even when beliefs are elicited, if beliefs and preferences are correlated, then this correlation may need to be modeled for correct inference. The way the researcher models this correlation has implications for achieving unbiased estimates.

To my knowledge, the only empirical paper in the literature that allows for a correlation of preferences and expectations about others' actions is Bellemare et al. (2008). However, their aim and approach in doing so are very different, as they are motivated by the need to control for a potential spurious correlation between the two to avoid an endogeneity bias. In particular, they addresses the identification of inequity averse preferences from choices in ultimatum games by collecting proposers' expectations of respondents' reactions. They distinguish between preferences when the person has more versus less payoff than the other and assume that respondents' and proposers' preferences regarding social allocations come from the same distribution. Distribution of preferences is estimated by a model that incorporates respondents' and proposers' actions in an ultimatum game and proposers' subjective expectations. The authors note the endogeneity bias in the estimates of preference parameters that may result from a potential spurious correlation of preferences and beliefs, for example, due to people with certain preferences also being more optimistic. Therefore, they allow for a direct correlation between the unobserved heterogeneity in preferences and the unobserved heterogeneity in beliefs. They find a strong correlation between

optimism and inequity aversion.

The results of my paper show that the relationship between preferences and beliefs produce differences in beliefs that are more than person-specific, and imply that the form of correlation modeled in Bellemare et al (2002) is incomplete, potentially failing to account for the simultaneity bias. Their assumption that correlation is spurious results in the modeling feature that a proposer's expectations about responders' choices are similarly optimistic, regardless of whether the responder has less or more payoff, as unobserved heterogeneity in beliefs are solely person-specific. However, my results imply, for example, that a proposer might be relatively more optimistic (compared to other proposers) when the responder has less payoff than him versus when he has more, if the person himself is not sensitive (while other proposers are) to having more or less. While significant person-specific correlations as Bellemare et al. (2008) report are a result of the relationship between preferences and beliefs I document in this paper, they are insufficient to account for it. In order to fully account for the correlation of preferences and beliefs, researchers should model their relationship at the individual- and game-specific level, by allowing beliefs to depend on how the individual values the options others are considering in that game.

Next, I provide an overview of the modified dictator games employed in this paper and discuss the model and estimation of altruistic preferences, as all the studies share these building blocks. In the following sections, I focus on the novel contributions of the paper. I detail the design and goal of each study, and explain the findings. I conclude with a discussion and implications for future research.

2 Review of Estimating the Degree of Concern for Others

2.1 Altruistic Preferences

In this paper, I define *altruism* as the non-strategic willingness to sacrifice one's own payoff due to a regard for another person's payoff. In a simple model, as Charness and Rabin (2002) proposed, the non-negative weight an individual puts on another person's payoff relative to his own can capture the degree of altruism. For people who are purely self-interested, this weight is zero. The variation in this relative weight can summarize differences in the regard for others. Moreover, people may be sensitive to relative position, positing far greater concern for those who have less than themselves

(behind) than those who have more (ahead). People may also be heterogeneous in the degree of concern withdrawal due to relative position. I describe the utility a person gets from a social allocation as a function of individual payoffs

$$u_i(x_i, x_j) = (1(x_i \geq x_j))[(1-\alpha_i)x_i + \alpha_i x_j] + (1(x_j > x_i))[(1-\beta_i)x_i + \beta_i x_j] \quad (1)$$

where individual i's utility is a weighted sum of her own payoff and the payoff of individual j. If a person is ahead, α captures her degree of altruism, and if she is behind, β captures it. If the person is purely self-interested then $\alpha_i = \beta_i = 0$. Although some people do not exhibit a positive regard for others, we may expect others sacrificing to help others, more so at cheaper helping prices, especially when they are ahead. Person-specific altruism parameters, α_i and β_i, can be inferred from individuals' choices in social allocation tasks that vary in the individual-sacrifice to social-benefit tradeoff. Experimental studies commonly use social allocation decisions in modified dictator games to measure individual differences in the degree of altruism (Charness and Rabin (2002), Fisman et al.(2007)).

2.2 Modified Dictator Games

In all the studies to follow, I infer an individual's degree of altruism from his choices across several modified dictator games. This section explains the structure of the modified dictator games used in this paper.

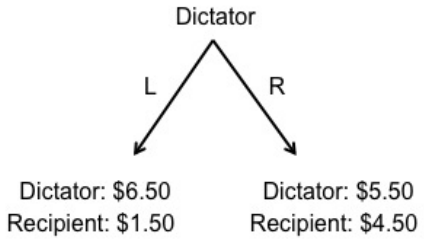
Each dictator game offers a binary choice between two social allocation options, (x_i^l, x_j^l) and (x_i^r, x_j^r). Person i makes a social allocation choice that determines his payoff x_i as well as the payoff of another person x_j. Person j (recipient) gets paid according to the person i's (dictator's) decision. The lack of the recipient's decision power makes this game a non-strategic choice task for the dictator.

In each game, one of the options pays the dictator more, but is socially inefficient. The other option involves a sacrifice from the dictator but yields a social benefit. Let the two social allocation options be denoted with subscripts L and R. For parsimony of exposition, assume option L pays the dictator more than option R, $x_i^l > x_i^r$. Then option R pays the recipient more than option L, $x_j^r > x_j^l$ and the dictator's sacrifice is smaller than the recipient's gain[2], $x_j^r - x_j^l > x_i^l - x_i^r$. So by

[2]In all the games in this paper, the dictator's sacrifice increases the social welfare since I am only interested in

picking the option that pays him less (R), the dictator can increase the payment to the recipient by more than the amount he sacrifices. Therefore, choice of this option is social welfare maximizing. In each game, either the dictator gets more than the recipient in both options, $x_i^l > x_j^l$ and $x_i^r > x_j^r$, or vice versa, $x_i^l < x_j^l$ and $x_i^r < x_j^r$. Figure 1 illustrates a sample game.

Figure 1: Example of a modified dictator game.



A participant makes decisions across multiple games that vary in the ratio of the dictator's potential sacrifice to the recipient's potential benefit. Given the choice between two options, person i will choose the option that pays him more (L) if the tradeoff between the sacrifice required and the increase in social welfare is larger than his willingness to sacrifice. If the dictator is ahead, $x_i^l > x_j^l$ and $x_i^r > x_j^r$, the preference formulation in Equation 1 implies he will choose L if $\frac{\alpha_i}{1-\alpha_i} > \frac{x_i^l - x_i^r}{x_j^r - x_j^l}$. If the dictator is behind, $x_i^l < x_j^l$ and $x_i^r < x_j^r$, he will choose L if $\frac{\beta_i}{1-\beta_i} > \frac{x_i^l - x_i^r}{x_j^r - x_j^l}$. Rearranging, if the degree of altruism in the given context, α_i or β_i, is smaller than a cutoff $Q \equiv \frac{x_{ik}^l - x_{ik}^r}{(x_{ik}^l - x_{ik}^r) + (x_{jk}^r - x_{jk}^l)}$, the person will choose the option that provides lower social welfare but higher personal payoff. By observing an individual's choices as Q changes, we can infer his degree of concern for others when ahead and when behind.

Tables 2, 3, 4, 9, and 14 list the set of modified dictator games subjects faced in each study and the percentage of the subjects who picked the option that paid them more (i.e., the selfish choice). These tables highlight general tendencies among the subjects regarding social allocation choices. As expected, subjects were more likely to be selfish when the options offered more to the recipient

eliciting heterogeneity in altruism and beliefs about others' willingness to make welfare-increasing sacrifices. If the goal were to determine the distribution of types of social preferences in the population as in Fisman et al. (2007) then determining whether a person who does not engage in any social-welfare increasing behavior would choose to engage in pareto-damaging behavior would be of interest. In that case, inclusion of pareto-damaging sacrifice options would help discern whether α or β are zero or negative. Given both the more specific objective of using variation in altruistic preferences as an application to study beliefs about others and the previous finding that most people do not engage in pareto-damaging actions (Charness and Rabin (2002)), I choose to focus on the degree of positive altruism. The inability to distinguish between whether a person is only self-interested or would sacrifice to damage social welfare only decreases the observed variation in preferences and potentially leads to an underestimation of the documented correlation between preferences and beliefs.

(when the dictator was behind) and when their sacrifice did not translate to a large increase in the recipient's allocation.

2.3 Inference of Altruism Parameters from Choices in Modified Dictator Games

I estimate a subject's degree of altruism from his choices in several dictator games assuming observed behavior corresponds to individuals implementing the preferences formulated in Equation 1 with error[3]. The decision maker is allowed to potentially choose randomly in some games and thus pick the utility-maximizing option with some probability less than one. Therefore, the error reflects a random mistake, and the person is equally likely to make a mistake in each dictator game. Then the likelihood of the individual picking option L over option R can be written as

$$Pr(y = l) = 1(u_i(x_i^l, x_j^l|\alpha_i, \beta_i) > u_i(x_i^r, x_j^r|\alpha_i, \beta_i))(1 - \omega_i) + \frac{\omega_i}{2}. \quad (2)$$

This approach rationalizes observed inconsistencies in the same way regardless of where the intransitivity occurs, and it may even result in set identification, where two preference parameters are equally likely. In a context with sparse discrete choice data, this approach is behaviorally and informationally more compelling than letting the likelihood of the error be a function of the utility cost of an error[4]. I apply this formulation to the choices made in dictator games to estimate subject specific α_i, β_i, and ω_i.

Appendix A2 displays individual-specific parameter estimates[5]. The reader will note the het-

[3]A simpler approach to characterizing differences in the degree of altruism across people is to rank them by the number of altruistic choices they made across the dictator games they faced, either as ahead or as behind. However, this approach does not distinguish between a person who is consistently middle of the road in his degree of altruism and a person who is just choosing randomly. In addition, this approach cannot predict what the person would choose given a novel dictator game. Therefore, I find it necessary and useful to work with a specific utility formulation.

[4]An example of modeling the error as a function of the utility cost is the logit regression. A unique preference parameter minimizes the individual's total error. In estimating population mean preference parameters for aggregate data, Charness and Rabin (2002) use this approach. Costa-Gomes, Crawford, and Broseta (2001), on the other hand, employ uniform errors, as in this application. The results of the paper are not sensitive to how the error term is modeled, and the replication of results using a logit formulation are available upon request.

[5]As each individual makes decisions in dictator games with two choice options, the preference parameters are identified up to the ranges that discrete questions divide the space. I can only determine whether an individual-specific preference parameter falls in the ranges of [0, .111], [.112, .143], [.144, .2],[.201, .25],[.251, .334],[.335, .4],[.401, .429],[.43, .5], depending on the set of questions asked. Therefore, to capture the observed heterogeneity among subjects, I assign parameters close to the minimum of these ranges, 0, .12, .15, .21, .26, .34, .41, .44, to avoid overestimating altruism. My later results about the relationship between preferences and beliefs are not sensitive to this choice. Moreover, I only use choices in modified dictator games that do not have zero or equal outcomes to estimate these preferences, as I find a sensitivity to these characteristics, yet do not want to complicate the utility model beyond what Charness and Rabin (2002) propose for this application. As long as estimated preferences can help infer how people evaluate options that others are considering, then this restriction does not influence my results.

erogeneity of preferences in the degree of regard for others when dictators are ahead or behind. The difference between the two parameters also varies greatly. Table 1 displays the joint frequency of α_i and β_i for those individuals who made decisions in both relative positions (ahead and behind) in Studies 1 and 3. We see that, in general, people show a withdrawal of concern for others when the game offers them less (behind) and others more, whereas the degree of the impact of relative position varies across people.

The specification of the altruistic utility function and the estimation of the individual altruism parameters α_i and β_i are in line with previous research. This paper does not aim to test between different models of social preferences and only focuses on positive altruism. Estimating the differences across people in their regard for others is the main goal, as this heterogeneity is essential to studying the relationship between individuals' preferences and beliefs about others' actions in similar contexts. The previous evidence on the heterogeneity in these preferences and the need to separate strategic kindness from true altruism in social interactions motivate the particular focus on altruistic preferences.

3 Studies and Results.

3.1 Study 1.

In this study, subjects first made choices in modified dictator games and then expressed their beliefs about other participants' choices. I designed the study to document the form of the correlation between beliefs about others' choices and the individual's own preferences that determine such choices. The main structure of the study allows me to draw a parallel with the findings in the false consensus literature. I also include new controls and an incentive-based belief elicitation method in order to bolster the generality of the findings. Most importantly, I augmented the design to be able to test new hypotheses regarding the nature of the correlation of interest.

Therefore, I also only study beliefs about others' choices in modified dictator games that do not have zero or equal outcomes.

3.1.1 Procedures and Design

One hundred and two undergraduate and graduate students participated as decision makers in this study[6] across three sessions of 24, 48, and 30. The experimenter informed the decision makers that the study had two parts and that their decisions in both would affect the payment they would receive in addition to the flat rate of $6 for their 30-minute participation. All payments were determined and made at the end of the experiment.

Decision makers read instructions detailing the nature of the decision tasks in the first part, how their choices determined both their payment and that of an anonymous recipient, that the computer would pick one decision task from the first part at random to determine payments, and how these payments would be made to ensure anonymity. The instructions also emphasized that decision makers were randomly and anonymously matched to a different recipient for each game, to eliminate potential dynamics in behaviors in a repeated interaction. These instructions, along with examples of the presentation of decision tasks can be found in Appendix A1.1. After participants read all the instructions, they were presented consecutively with several modified dictator games after the instructions. In each game, they had to choose between two social-allocation options. The games varied in the ratio of the dictator's potential sacrifice to the recipient's potential benefit. Some subjects made decisions only in games that paid them more than the recipient in both social allocation options, whereas some also faced games that paid them less than the recipient in both social allocation options.

After decision makers completed the first part, additional instructions informed them about the nature of questions in the second part and that their additional payment included a reward based on the accuracy of the subject's estimate of others' actions in the randomly picked question from this part. For each dictator game in this part, participants predicted the likelihood of each option being picked by the dictators in the experiment. Some subjects only made guesses about others' behavior in games they had faced themselves as decision tasks in the first part, some subjects only made guesses about games they have not faced before, and some made guesses about a mixture of

[6] Another 72 students participated as recipients in session 1 and 2 and had no decisions to make regarding interactions with dictators. In sessions 1 and 2, a coin toss determined whether the participant was assigned the role of the dictator. Participants in these sessions who did not make decisions in modified dictator games spent their time reading instructions about the specifics of dictator games and then answering an unrelated survey for 30 minutes in another room before receiving a flat fee for their participation along with the amount that a randomly matched dictator allocated to them. In session 3, all participants were in the same room and played alternating roles.

both.

The study consisted of three different sessions sharing the same basic design with different robustness controls. Appendix A1.1 includes all the procedures and design details. Sessions varied in the form of accuracy incentives used, whether the recipients were in the same room, whether the total population of other decision makers was visible, whether the relative position of the dictator was varied, and whether the expectations were measured as percentages or frequencies. Some sessions counterbalanced the order of options in modified dictator games and the order of questions. None of these robustness controls had any impact on the results, supporting the generality of the findings.

3.1.2 Preliminary Results

In Tables 2-4, I document the dictator games in all three sessions of Study 1 for which subjects were asked to state their beliefs. The tables include information on the percentage of the subjects that picked the selfish option over the social-welfare-maximizing option and the average beliefs regarding the percentage of others who would pick the selfish option. Tables 2 and 3 also compare average beliefs across subjects who made different choices in a game[7].

From these averages, we can identify three main patterns. The first is that subjects correctly recognize that fewer people are selfish when the tradeoff of sacrifice to social benefit is low. The second is a considerable amount of conservatism; beliefs are biased toward the uniform prior. The third is that substantial differences in subject's beliefs exist, depending on the subject's behavior in that game ($\beta = 0.33$, $t = 6.87$). In fact, the third pattern is congruent with what the false consensus literature has documented across many domains (for a review, see Marks and Miller (1987)). Direct elicitation of beliefs in Study 1 allows me to draw this parallel between my data and what the false consensus literature has documented. In addition, I document this pattern with monetary incentives for accuracy. Iriberri and Rey-Biel (2009) independently provide additional evidence for the correlation of dictators' choices and beliefs about other dictators' actions using accuracy incentives. While the first study in that working paper and Study 1 in this paper elicit

[7]In cases where the subjects had made a choice in that dictator game in part 1, their actual choices from part 1 help split the data. If the subjects had not seen the games then their predicted choices are used to split the data. I predict their choices by using their invidual preference parameters α and β estimated from their choices in part 1. In session 1, all of the belief questions were presented in Part 1, whereas in session 2, only some were. In session 3, none of the beliefs questions were presented in Part 1.

similar data from subjects, the two papers diverge in their research question, analysis and other studies.

3.1.3 Discussion of Results and Motivation for Further Analyses

These results provide additional support for systematic differences in beliefs as a function of own choice, which is a clear violation of the assumption of mutual consistency. As previously debated (Hoch 1987, Dawes 1989, Krueger and Clement 1994), this systematic correlation may result from egocentrism or from informational differences across subjects. Whatever its origin, such systematic differences in beliefs stand in the way of inferring preference from choice under uncertainty. For example, if relatively selfish dictators believe more of the population is selfish than altruistic dictators do, their willingness to defer to others in a trust-dictator game that involves social allocations (as I study in Study 3) will be low, not only because they are less altruistic, but also because they don't trust others to be altruistic. Therefore, we cannot ascribe differences across people in their behavior to differences in their preferences alone.

These systematic belief differences may further complicate the econometric inference problem. If beliefs are not orthogonal to choices (nor, as I will show, to preferences) then not only would the econometrician want to collect data on beliefs, but she might also need to model this correlation appropriately to avoid biases resulting from endogeneity (Bellemare et al. (2008)). In order to understand the nature of this correlation with the aim of incorporating it into an appropriate econometric model, I ask the following three questions: (1) Does the act of having made a choice matter? In many situations, people form expectations about others' decisions without having faced the exact decision themselves. If the person (and the econometrician) can predict what he *would* choose in that situation, does this predicted action influence beliefs in the same way? (2) Do differences in subjects' choices fully capture the systematic differences in beliefs? Consider two people with the same observed choice in a given dictator game. According to the previous findings in the false consensus literature, their opinions should be similarly skewed toward expecting a higher percentage of others to make this same choice. However, two people making the same choice could have different preference parameters. As a result, one could have a small utility difference between the two options in the dictator game, whereas the other could have a large one. If preferences instead of choices alone affect beliefs, the person who is almost indifferent between

the options and the person who strongly favors his choice may have systematically different beliefs about others' choices in the same context. (3) Are differences in beliefs person-specific alone? A person might be pessimistic about others' altruism regardless of how he evaluates the options in the given context, which means the person's relative pessimism is person-specific[8]. Alternatively, the relative pessimism of the person can depend on his evaluations and therefore potentially change across contexts.

To answer these questions, I move the focus of the analysis from choices to the underlying preferences to capture the heterogeneity in beliefs. I make use of the individual specific α_i and β_i, estimated from choices in modified dictator games[9], both in constructing what the subject would most likely choose in a new game and in capturing differences of preference intensity among subjects with the same discrete choice. Using individual preference estimates to quantify the differences in option evaluations across contexts for a given person helps answer whether the relationship between preferences and beliefs are person-specific alone.

3.1.4 Model and Empirical Analysis

I model an individual's beliefs about others' choices to vary with how he evaluates the choice others face. To capture the false consensus effect I documented above, I allow one's choice to affect her beliefs about others' choices in that same game. I also allow beliefs about actions in a game the individual has not faced to potentially depend on what she would choose in that game, given her estimated preferences. Over and beyond the effect of actual or potential choice, I also consider additional heterogeneity due to a finer classification of individuals, by their preference. To this end, I allow for the potential of individuals' guesses about others' choice probabilities to depend on how far from making each particular choice they see themselves.

The resulting basic model of person i's beliefs about the proportion of people picking the selfish option (say L) in dictator game k is

[8]This outcome may potentially result from a spurious correlation, such as selfish people being more pessimistic in general.

[9]When the parameters can take on more than one value, I use the smallest. None of the results are sensitive to this choice; robustness checks are available upon request.

$$
\begin{aligned}
B^*_{ik} &= \mu_k + \varepsilon_{ik} + 1(x^{l,r}_{ik} > x^{l,r}_{jk})[1(\alpha_i < Q_k)\rho + (Q_k - \alpha_i)\delta] \\
&+ 1(x^{l,r}_{ik} < x^{l,r}_{jk})[1(\beta_i < Q_k)\rho + (Q_k - \beta_i)\delta]
\end{aligned}
$$

B^*_{ik} is person i's expectation of the percentage of others who choose L in game k. Q stands for the cutoff $\frac{x^l_{ik} - x^r_{ik}}{(x^l_{ik} - x^r_{ik}) + (x^r_{jk} - x^l_{jk})}$. Technically, $B^*_{ik} = Prob^i(\alpha_{-i} < Q_k)$ if $x^{l,r}_{ik} > x^{l,r}_{jk}$, and $B^*_{ik} = Prob^i((\beta_{-i} < Q_k))$ if $x^{l,r}_{ik} < x^{l,r}_{jk}$.

The game-specific means, μ_k, capture variations in expected probabilities across games, such as the fact that the expectation of observing an altruistic choice should decrease with the sacrifice required, as well as other factors that deviate from objective probabilities, such as general pessimism or conservatism[10]. Given these game-specific controls and an i.i.d. error term ε_{ik}, the assumption of mutual consistency rules out further differences.

The remaining terms allow for additional systematic differences as a function of the individual's preferences. In games where the dictator is ahead, $x^{l,r}_{ik} > x^{l,r}_{jk}$, the altruism parameter α_i determines evaluations of options and consequently the choice. If the degree of altruism, α_i, is less than the cutoff, Q_k, then the individual chooses the selfish option providing lower social welfare. Similarly, if the dictator is behind, β_i governs evaluations of options. Therefore, given estimates of α_i and β_i, even if the subject has not made a choice in game k himself, we can determine which choice person i's preferences would lead him to.

The effect of one's own discrete choice (observed or predicted) in game k on one's belief about others' choices in game k is captured by ρ. The parameter δ, on the other hand, investigates further systematic belief differences among individuals with the same choice but different preferences. The larger the difference between the preference parameter and the cutoff, the larger the utility difference between choosing the selfish versus the altruistic option. If people assume a similarity between their underlying preferences that shaped their choice and the preferences of others, then we would expect that between two people who made the same choice, the person with the larger utility difference between options is more likely to believe more people chose the option he prefers. Therefore, we

[10]Please see Appendix A3.1 for results on mean-beliefs inaccuracy results. I find considerable conservatism but no support for a common pessimism about altruism in the population. I also discuss the impact of systematic heterogeneity in beliefs on the measurement and interpretation of common inaccuracy effects such as conservatism and pessimism.

expect δ to be positive.

While estimating this basic model, I take into account the fact that reported beliefs are between zero and one with a censored regression.

$$B_{ik} = \begin{cases} 0 & \text{if } B^*_{ik} < 0; \\ B^*_{ik} & \text{if } 1 \geq B^*_{ik} \geq 0; \\ 1 & \text{if } B^*_{ik} > 1. \end{cases}$$

The assumption that the errors ε_{ik} are distributed normally produces a two-limit tobit model, guaranteeing the reported beliefs are between 0 and 1. I treat the preferences parameters α_i and β_i as data in the estimation of beliefs.

In the empirical analysis, I extend this basic model to allow for further interactions of interest. For example, I observe two individual characteristics, gender and area of study. I use these to test for differences in mean beliefs and in the degree of preference and belief correlation subjects display. Since economics majors should have more experience with modified dictator games, they should display lower levels of correlation between own choice and expectations of others' choices, per the informational argument of Dawes (1989). They may also be more pessimistic on average about altruism in the population. I do not have any reason to expect gender to be a significant factor.

Another person-specific factor that may influence the degree of correlation between preferences and beliefs is ω_i, obtained from the estimation of individual preference parameters. The more the number of choices the subject made that were irreconcilable with the underlying preference parameter estimated, the higher this parameter. I do not expect this factor to influence mean beliefs. The effect of this parameter on the degree of correlation between preferences and beliefs is unclear. If we interpret ω_i as capturing carelessness, take the view that the relationship between preferences and beliefs result from an egocentric bias (projection), and assume careful deliberation leads to seeing different ways others can react, we would expect a positive effect of ω_i, as we expect those who are more careful to project less. On the other hand, if we interpret ω_i as diffuseness of preferences then we can expect people with more certain preferences to have a stronger correlation between their preferences and what they think others will do.

Finally, I allow the act of having made a choice to influence the correlation between preferences

and beliefs. If having made a choice in a decision task makes one's choice more salient then we expect the correlation between beliefs and these actions to be stronger. I need to use within-person variation in whether the participant faced the task in Part 1 to identify this effect. Therefore, only session 2 provides suitable data to test the hypothesis.

3.1.5 Results

Given individual preference estimates, I can predict subjects' choices in new games. Tables 3 and 4 compare average beliefs across subjects who sare predicted to make different choices in a game, based on their α_i's and β_i's. The raw data present a similar pattern of belief differences based on predicted choices compared with the differences based on actual choices.

Table 5 presents the results for the tobit regression of beliefs using data from all three sessions[11]. The differences across subjects' own choices in the raw data suggest that considerable heterogeneity exists in beliefs about others' propensity to pick the selfish option. Column 1 documents that, on average, a subject who picks the selfish option estimates the percentage of others who would pick the selfish option to be 28.4 percent higher than would a subject who picks the altruistic option. Table 6 displays the results of this first regression across each session in the study. The results across these sessions suggest that this evidence for projection is not sensitive to the robustness check differences[12].

Column 2 in Table 5 documents the effect of underlying preferences on beliefs over and beyond the effect of one's own choice. The larger the difference between Q (the cutoff in the question) and the degree of one's altruism, the further the subject's evaluations are from making an altruistic choice, and I find the higher the estimated percentage of selfish actions. Importantly, this result provides the first evidence in the literature that differences in choice do not capture all the systematic variation in beliefs. Differences in the strength of preference introduce further heterogeneity.

If beliefs vary with preferences and not just choice, the systematic heterogeneity has a bigger

[11] In all the analyses, I drop subjects whose parameters are estimated with ω_i (propensity to choose randomly in dictator games) more than 40% in order to have reliable predictions about the options they would have picked in new games given their preferences. As a result, I drop 15 out of 102 participants (3 subjects from session 1, 11 from session 2, and 1 from session 3) while preserving a considerable amount of variation in ω_i. I also do not include beliefs about choices in modified dictator games with zero or equal outcomes to preserve consistency across sessions in this study and in other studies, and to keep the utility model simple, as aversion to zero is another heterogenous preference parameter. My results are not sensitive to any of these exclusions and robustness checks are available upon request.

[12] All other findings are also congruent across sessions and replication is avoided for brevity.

individual-specific component. However, note that the above proposed model does not imply that all heterogeneity is individual-specific; the person's relatively high beliefs about the selfish option being chosen, compared to objective probabilities, decrease for questions in which she is more likely to make an altruistic choice. For example, if a person has a relatively high α_i but a relatively low β_i then in games where the dictator is ahead, she will be less pessimistic compared to others than in games where the dictator is behind.

Note, however, that a significant ρ in the first regression can well be a result of individual-specific differences in pessimism about others' altruism regardless of how individuals evaluate the options in a given context. This outcome may potentially result from a spurious correlation, such as selfish people being more pessimistic in general. In fact, the canonical experimental false consensus studies are subject to this concern, as they usually only ask one belief question per person regarding the same context. However, I ask subjects to report expectations across many games, and therefore I can test this hypothesis by including subject fixed effects to control for individuals' mean level of deviation from the population average beliefs, μ_k's, across games.

Columns 3 and 4 of Table 5 document that these systematic individual differences are not merely individual differences in pessimism across people, as the effect remains significant even after subject fixed effects are included. This result suggests person i's relative deviation from game-specific belief averages μ_k changes as his evaluation of the options change. This finding is at odds with the specification Bellemare et al. (2008) use to control for the correlation of beliefs and preferences. They model differences in beliefs to be person-specific and not to have a decision-context-specific component. Yet the current results show that a direct correlation of preferences and the level of individual-specific pessimism will not suffice.

Given that subjects do not display the same level of optimism or pessimism across different modified dictator games in general, I now test whether people who have high concern withdrawal due to relative position (large $\alpha_i - \beta_i$ difference) also think that others will display high concern withdrawal[13] . Regardless of a people's individual tendencies to be pessimistic or optimistic, if they project their concern withdrawal onto others, we expect those with higher levels of concern withdrawal to display a bigger difference in their beliefs of altruism of other dictators in games where the dictator is ahead versus behind. Table 7 reports the results of a tobit regression of

[13]Only subjects in session 3 faced dictator games in both positions.

beliefs on the interaction of whether the dictator is ahead in the dictator game and the $\alpha_i - \beta_i$ difference of the subject, controlling for subject specific and game specific averages. To expose the general differences in beliefs due to the relative position of the dictator and of those who have more sensitivity to relative position, column 2 displays the results of a regression without subject and game controls. We see that the bigger the concern withdrawal of the subject, the more he thinks that the dictators will be less selfish when ahead than when behind. In sum, within-person differences in beliefs reflect that subjects think others are similar to themselves in the sensitivity to relative position.

Turning to testing for differences in the degree of the correlation between preferences and beliefs, I first test if whether the subjects made a choice in the decision task before increases the correlation of preferences and beliefs. Session 2 provides within-subject variation in this regard, as subjects see a mixture of new tasks and tasks in which they have made choices. I include a mean control for the questions the subjects had faced, along with the necessary interactions as follows:

$$\begin{aligned} B^*_{ik} &= 1(x^{l,r}_{ik} > x^{l,r}_{jk})[1(\alpha_i < Q_k)\rho(1+\nu_1 1(new_{ik}=0)) + (Q_k - \alpha_i)\delta(1+\nu_2 1(new_{ik}=0))] \\ &+ 1(x^{l,r}_{ik} < x^{l,r}_{jk})[1(\beta_i < Q_k)\rho(1+\nu_1 1(new_{ik}=0)) + (Q_k - \beta_i)\delta(1+\nu_2 1(new_{ik}=0))] \\ &+ \nu_3 1(new_{ik}=0) + \mu_k + \varepsilon_{ik} \end{aligned}$$

where $1(new_{ik}=0)$ indicates person i had faced game k as a decision task in Part 1.

Column 1 in Table 8 shows that subjects do not display a higher or a lower correlation between their choices and their beliefs if they made a choice in the task before as a dictator. Neither do they have different mean beliefs. Column 2 in this table replicates column 2 of Table 5 for session 2, and column 3 shows no impact of having faced the task on the relationship between beliefs and preference intensity. Therefore, I conclude that subjects' predicted choices relate to their beliefs in the same way their actual choices would. This evidence suggests subjects' commitment (or absence thereof) to a specific choice has no impact on their beliefs about the choices of others, which means researchers do not have to collect choice data in the same contexts as the ones in which beliefs are elicited; knowledge of the relevant individual preference estimates would suffice.

I also test whether person-specific observables such as gender, area of study, or error-proneness

in modified dictator choice tasks (ω_i) capture any heterogeneity in the level of correlation between preferences and beliefs. Gender and area of study do not have any effect on the degree of correlation between beliefs and preferences. However, I find that the people who are more consistent in their choices in Part 1 project these choices more strongly onto others. Table 9 demonstrates this finding, providing evidence against the hypothesis that increased deliberation decreases projection.

These results provide evidence for a correlation structure between preferences and beliefs that extends the correlation of choices and beliefs, and that adds to allowing for mere differences across subjects. Being aware of not just the existence but also the structure of this correlation informs both the type of data researchers need to collect to identify preferences and the type of model required to relate the data.

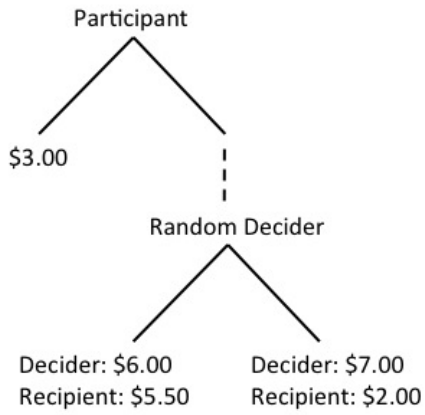
3.2 Study 2.

The goal of Study 1 was to document the existence and form of the correlation between preferences and beliefs. To this end, I elicited beliefs about others' choices in different decision contexts, consistent with the previous literature on false consensus and expectation measurement (Manski 2004).

The aim of Study 2 is to present additional evidence on the nature of the correlation between one's degree of altruism and one's beliefs about others' altruism without directly asking them about their beliefs. I do so by relating participants' altruistic preferences to their risk-taking behavior in gambles on others' altruism. As in Study 1, I inferred individuals' altruistic preferences from their actions in modified dictator games; however, instead of guessing about others' behaviors, subjects made risky decisions involving their beliefs about others' behaviors. These decisions involved a choice between a sure payoff and a risky option that paid what a random previous dictator allocated to a recipient in a given modified dictator game. Figure 2 illustrates an example of such a decision task.

The participants risked earning less money than the sure payoff if the dictator chose the option that paid the dictator more, and they stood to earn more money if the dictator chose altruistically instead. Therefore, the higher the subject's expectations are about altruistic behavior in the given game, the greater his expected utility from taking the gamble. Subjects' decisions only influenced their own payoff, and therefore concern for others did not influence their decisions. However, given

Figure 2: Example of a gamble on others' altruism.



the evidence of the relationship of beliefs about others' altruistic actions and own preferences from Study 1, I also expected systematic differences in the willingness to take a gamble in which these beliefs play a role.

3.2.1 Design and Procedures

Departing from Study 1, Study 2 controlled for the sequence of elicitation. It consisted of two seemingly independent experiments run two weeks apart with the same panel of subjects. The order of the experiments was counterbalanced. Among the 51 subjects who took part in both experiments, 23 participated in Experiment B first, in which they made decisions involving gambles on altruism, and 28 participated in Experiment A first, in which they made decisions in modified dictator games. I allowed only those students who had not participated in Study 1 to participate in this study.

In Experiment A, subjects were presented with 16 modified dictator games drawn from those presented in Study 1, Session 3. The games were divided into two different sets of eight. Each subject played the role of the dictator for one set of games and played the role of the recipient for the other set. Games from each set were interspersed. Each set included four games where the dictator was behind and four where he was ahead. For each relative position, the sets presented the following ratios between the dictator's sacrifice to the recipient's benefit: {1/8, 1/6, 1/3, 1/2}. Games presented the payments in tokens, where 100 tokens equaled $1. Table 10 presents these games and average behavior statistics.

At the end of Experiment A, the experimenter picked one modified dictator game at random to finalize payments. Subjects were paid according to their role in that game. If they were a dictator, their decision determined their payment. If they were a recipient, the dictator's decision determined their payment. Each dictator was randomly and anonymously matched to a different recipient for each game. Experiment A paid \$16.17 on average for participation and the additional payments resulting from decisions made in the experiment.

Experiment B consisted of three parts. Part 1 asked subjects to first make risky decisions involving picking between a sure outcome and a gamble with known probabilities, where the sure outcome was titrated between the minimum and the maximum payoff of the gamble in \$0.25 increments. The gambles were [33%, \$6.00; 67%, \$2.00], [60%, \$7.50; 40%, \$3.50], [75%, \$4.50; 25%, \$1.50] and [20%, \$6.50; 80%, \$2.00]. After Part 1 was collected, Part 2 presented four risky choices between a sure payoff and one of the gambles above, to check for consistency. Part 2 continued with detailed explanations and instructions about decisions involving a sure payoff and a gamble on a random past participant's decision in a modified dictator game. Appendix A1.2 contains these instructions and further details.

After reading the instructions, subjects made risky decisions involving picking between a sure outcome and what a past dictator allocated to a recipient in the particular modified dictator game, as Figure 2 depicts. The participants were told that the past experiment took place in the same lab with a random selection of students from the same subject pool. The instructions informed the participants that the dictator had chosen an option in the past without knowing the existence of the decision task the participant faced today. Moreover, the instructions clarified that the participants' decision between the sure payoff and the risky option would not influence the payoffs of any participants in the current or the past experiment. Thus their decisions depended on their subjective beliefs about dictators' choices in the specific modified dictator game and not on their degree of concern for others.

Experiment B included four dictator games (two ahead / two behind) from each set of eight in Experiment A, ensuring that each participant had seen (or was going to see) all the games in Experiment B but acted (or was going to act) as a dictator in only half of them in Experiment A. For each of the eight modified dictator games in Experiment B, several risky decision tasks were presented where the sure outcome was titrated between the minimum and the maximum payoff the

dictator could give to the recipient in $0.25 increments. After Part 2 was collected, Part 3 presented the same modified dictator games in a different order and asked the subjects to choose between a sure outcome and what the dictator allocated to the recipient in that game. Table 10 displays the average certainty equivalents and entry behavior for the games presented in Experiment B.

The consent form informed subjects they would receive $6.00 for their participation at the end of the experiment and an additional payment between $1.50 and $7.50 depending on their decisions and the outcome of the randomly picked gamble. If the subject chose the gamble over the sure payment, the objective probability of each event and a random number generator determined the outcome for each individual. The percentage of choices of dictators who participated in Study 1, Session 3 and a random number generator determined the outcome for gambles involving modified dictator games. The average payment was $10.47.

3.2.2 Analyses and Results

Given subjects' risky choices across the titration of sure payoffs in $0.25 increments for each gamble, I can infer the certainty equivalent for that gamble for any individual (up to the increments) using the point where the subject switches his choice from the sure payoff to the gamble. Summary statistics in Table 10 include average behavior in modified dictator games, average certainty equivalents for gambles based on these games[14], and frequency of choosing to gamble at a given sure payoff. Average entry behavior is consistent with the average certainty equivalents (CE). If the sure payoff is above the average CE then entry is lower; if the sure payoff is below the average CE then entry is higher. When the sure payoff is close to the average CE about half the subjects enter.

Note that if the dictator chooses selfishly in the game presented in the gamble, the subject gets less than if he would have chosen the sure payoff. Conversely, if the dictator sacrifices for the benefit of the recipient, then the subject gets more than the sure payoff. Therefore, the higher the subject's expectations are about altruistic behavior in the given game, the greater his expected utility from taking the gamble. Therefore, each gamble presents a risky decision task in which the subject needs to make a decision based on his level of trust in others' altruism in the particular modified dictator game.

[14]I take the certainty equivalent to be the minimum of the $0.25 range for the calculations in this table.

Given subjects' own choices in modified dictator games in Experiment A, I either observe what the subject personally would choose in the game presented in the gamble, or I can predict his choice based on his estimated preferences, as in Study 1. The last two columns of Table 10 summarize subjects' entry behavior and CEs based on their observed or implied personal choices in the shoes of the dictator.

We see the subject is more likely to choose the gamble over the sure payoff if he would have chosen the welfare-maximizing option had he been in the shoes of the dictator. If he would have chosen the selfish option, he is more likely to avoid the gamble and opt for the sure payoff. We also see a parallel pattern in certainty equivalents. In gambles where the objective risk is moderate, the difference in certainty equivalents between the two groups is more pronounced. However, the raw data in certainty equivalents are less pronounced because when a gamble has a high expected value due to low risk or a low expected value due to high risk, the difference across groups is depressed due to truncation at one end.

By design, a subject's choice between the gamble and the sure payoff does not affect any other subject's payoff. However, Table 10 suggests that subjects who are more concerned about others are more likely to pick the gamble over the sure payoff. Therefore, we must conclude that either those with more concern for others place a higher expected value on the gamble or that they like taking on such gambles more than those with less concern for others. If risk preferences and the degree of concern for others are not orthogonal then one cannot conclude that systematic differences in entry decisions are due to systematic differences in beliefs that relate to differences in preferences.

To test for a correlation between social and risk preferences, I relate subjects' choices in risky decisions with known probabilities (Experiment B, Parts 1 and 2) to their choices in modified dictator games. I describe each individual with two variables, his level of altruism when ahead and when behind. Using a reduced form approach to avoid having to assume a specific form of risk aversion, I test whether risk-taking behavior correlates with these variables in two different ways. First, using the choices from Part 2, where the subjects were presented with only one sure payoff for each of the four gambles, I examine whether a subject's propensity to take the gamble correlates with his altruism variables. Table 11 presents estimates from a probit model of binary choice. Entry implies the subject took the gamble over the sure payoff. I model this decision to depend on the degree of altruism when ahead, the degree of altruism when behind, and fixed effects

for the different gambles. I do not find any evidence that differences in altruism explain differences in risk-taking behavior. The second test uses entry decisions in the titration task for each gamble imply (Experiment B, Part 1). Bearing in mind that the switching behavior (between choosing the gamble versus the sure payoff) can only happen at a discrete number of titrations and that these are ordered in value, I model the point at which the subject will switch (i.e. his CE) based on his degree of altruism when ahead, his degree of altruism when behind, and fixed effects for the different gambles, using an ordered probit. Table 12 presents the results. Differences in altruism do not correlate with differences in certainty equivalents[15].

Given that I find no evidence for a correlation of risk and social preferences, the inference is greatly simplified, as the knowledge that these preferences are orthogonal allows me to study risk-taking behavior in gambles on altruism without having to incorporate individual risk preferences[16]. I use the decisions participants made in Part 3 of Experiment B and a reduced form approach to model the propensity of taking a gamble on altruism over a sure payoff. Each subject made eight decisions between a sure payoff and a gamble on altruism. Each gamble involved a different modified dictator game. I use binary probit to model the decision to take the gamble as depending on one's own actual or inferred choice in the related dictator game. Column 1 of Table 13 presents the results. The coefficient -0.51 translates to a marginal effect of 20% less propensity to take the gamble if the participants' social preferences would have led them to pick the selfish option had they been in the shoes of the dictator. As in Study 1, I also examine the impact of preference intensity, measured by the difference between the tradeoff Q_k of the dictator game in the gamble and the subject's altruism parameter α_i or β_i, depending on the relative position of the dictator. The larger the difference is, the higher the utility of the selfish option over the altruistic option for the subject. Therefore, I expect preference intensity for the selfish option to decrease the propensity to enter, as shown in column 2 of Table 13. However, since this study does not employ as wide a range of tradeoffs for modified dictator games as Study 1, the results are significant but weaker due to the lack of evidence of wide variation in the strength of preferences across subjects.

We can also ask whether the differences in gamble-taking behavior are due to other subject-

[15] In this regression, the certainty equivalents are discrete choices and only the difference matter rather than which point in the range is picked as the specific certainty equivalent.

[16] An econometric model with a specific form of risk aversion can be estimated in order to incorporate individual differences in risk preferences. Capturing the variation due to risk would decrease the noise and produce more precise estimates for the impact of social preferences on implicit beliefs about others' altruism. However, one must weigh this benefit against having to rely on a specific risk-aversion formulation.

specific unobservables. Some examples of such differences could be differences in risk aversion or differences in general pessimism. Since we observe decisions of gamble-taking behavior within and across subjects, we can add subject-specific controls. I report the results of including subject fixed effects in column 3 of Table 13. I drop six subjects' decisions, as the subjects do not vary over the eight gambles in their decision to take the gamble. We see that all of the evidence for the correlation between preferences and gamble-taking behavior comes from within-subject variations.

Note that this study alternated the sequence of preference elicitation and risky choices. Table 14 reports two tests of sensitivity. Columns 1 and 3 include the results of the main analysis (Table 13, column 1) repeated separately for the two conditions. In one condition, subjects participated in Experiment A and then came back two weeks later to participate in Experiment B. In the other condition, the sequence was reversed. We see the correlation between one's choice and one's beliefs do not depend on whether he made choices in modified dictator games first or last. In the condition where subjects participated in Experiment A first, we can ask whether having made a decision in a previous dictator game changed the way in which own preferences affected beliefs. Column 2 of Table 14 reports the results. As in Study 1, we see no such effect. For completeness, I repeat this analysis for the other condition and report it in column 4. The fact that the subject faced a gamble regarding a dictator game does not influence decisions in that game.

3.2.3 Summary and Discussion of Study 1 and Study 2 Results

The first two studies presented evidence that differences in beliefs elicited with accuracy incentives, as well as differences in implicit beliefs operant in risky choices involving trust in others' altruism, are systematically related to differences in preferences. These systematic differences across people violate the commonly held assumption of mutual consistency when inferring preferences from choices made under uncertainty. Therefore, a careful documentation of the existence of these differences, with robustness checks relating to the standards of several disciplines, is essential before we study the impact of such differences in a context of strategic interactions. I show the systematic belief differences are robust to monetary incentives for accuracy, do not depend on whether beliefs and preferences are elicited first, and are not driven by people who are making more errors in the experiment or by the direct elicitation method.

However, more importantly, both studies further contribute to our understanding of the rela-

tionship between preferences and beliefs by ruling out previous empirical operationalizations. I show, for the first time, that these differences do not depend on differences in choices alone. People think others have underlying preferences similar to their own. Two people making the same choice have systematically different beliefs about others' choices, depending on the utility difference they perceived between the two options the choice presented.

Another important finding is that relative differences in the deviations of beliefs are not person-specific. As a person's utility difference of the choice options varies across decision tasks, so does his relative pessimism or optimism compared with the population. Therefore, a simple correlation of preferences and these beliefs, as in Bellemare et al. (2008), cannot account for the form of the relationship between preferences and beliefs about others' choices across different decision tasks. The relationship should be modeled at the individual- and game-specific level by allowing beliefs to depend on how the individual values the options others are considering in that game.

Consequently, the documented relationship between beliefs and preferences has particular implications for the required variation in belief data for preference inference from choices in strategic contexts. First, beliefs should be elicited for several contexts over which the individual's valuations for the choice alternatives vary. Second, beliefs should span a broad range of objective probabilities. If the range of probabilities is mainly restricted to one part of the [0,1] interval, potential conservatism will be confounded with pessimism and/or systematic heterogeneity in beliefs.

In Study 1, I use a reduced form approach to study within-person belief differences across games that present the dictator different relative positions. I find that the belief differences within a person who has a high sensitivity to relative standing (large α-β difference) are larger than those of a person with a lower sensitivity to relative standing. This finding means that regardless of how altruistic people think others generally are, those who have large concern withdrawal when they get less than the recipient think others display a higher sensitivity to relative position than those who have less pronounced concern withdrawal do.

I also show that the act of having made a choice does not affect the impact it has on beliefs, and subjects' predicted choices relate to their beliefs in the same way their actual choices would. This evidence suggests subjects' commitment (or lack thereof) to a choice does not change their beliefs about the decision task others face, which means researchers do not have to collect choice data in the contexts in which beliefs are invoked; knowledge of estimates of the relevant preference

parameters would suffice.

Given the evidence on the existence of systematic belief heterogeneity that violates the assumption of mutual consistency and the type of relationship between preferences and beliefs that lead to it, I proceed to studying strategic actions that depend on both one's preferences as well as his beliefs. My goal is to study the impact of this systematic belief heterogeneity on strategic actions, and the resulting consequences for preference identification from these actions.

3.3 Study 3.

Decision makers take two factors into account when choosing the optimal action in any strategic interaction: their evaluation of potential outcomes and the likelihood that each outcome will be reached given that action. Beliefs about others' actions shape these subjective probabilities. In Study 3, I present a strategic game where subjects' actions are based on their expectation of others' actions in modified dictator games, as in Study 2. However, in addition to their beliefs, their own preferences for altruism also guide their decisions. Thus, this game provides an example of how beliefs and preferences found in strategic actions codetermine choice. First, I present evidence of systematic differences in beliefs leading to discernable patterns in strategic actions. Then I demonstrate the misestimation of preferences if the systematic heterogeneity in beliefs is ignored.

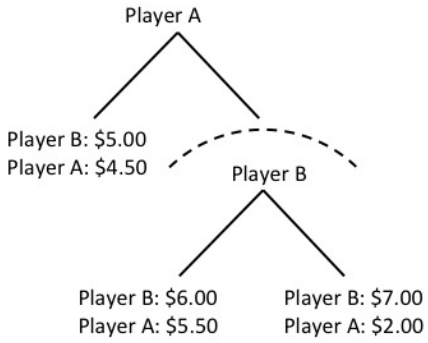
3.3.1 Procedures and Design

The third study involved 64 subjects to whom I presented a strategic social interaction in which the outcome of a two-stage game depended on the choices of two players. The first player (Player A) decided between taking a certain social allocation that determined both players payments or deferring the choice to Player B. Player B, on the other hand, faced a dictator game and reported her strategy without learning about Player A's choice or the certain allocation Player A was considering[17]. Therefore, Player B's decision in the sub-game depends only on Player B's altruism for an anonymous other person. Figure 3 illustrates this game:

Player A's entry (deferral to Player B), along with the decision of the matched Player B's in the dictator sub-game jointly determine the outcome for both subjects. This game presents a scenario in which Player A's decision to defer to Player B not only depends on his beliefs about Player

[17]I used this limited information scenario to minimize positive reciprocity concerns, even though Charness and Rabin(2002) and Gneezy (2000) fail to find any.

Figure 3: Example of a trust-dictator game.



B's choice but also on Player A's altruism. Therefore, it highlights the identification problem that arises in disentangling heterogeneity of preferences from the heterogeneity in beliefs.

This two-stage game can be thought of as a trust/dictator game. Charness and Rabin (2002) used this variant in characterizing social preferences without reciprocity. This game is also similar in design to the wallet game in Dufwenberg and Gneezy (2000); however, it is importantly differentiated by the information restriction that Player B only observes the sub-game.

A total of 64 subjects made 16 decisions in each of the two roles. Half of the subjects were first in the role of Player A and then played dictator games (i.e. in the role of Player B). The other half had the reverse sequence. The choice of deferral always involved a strict increase in payoffs for Player B; however, subjects were not informed of this[18]. In half of the games, Player A was ahead in all outcomes; in the other half of the games, Player B was ahead in all outcomes. At the end of the experiment, the experimenter picked one game at random by a toss of several dice and made payments by matching subjects anonymously according to a predetermined list that ensured each subject was matched with a different person in the other role for each game. While this matching took place, the last part of the experiment asked subjects to state their beliefs about the percentage of other subjects who picked each option across several dictator games. The experiment provided no monetary accuracy incentives for these beliefs, as the simultaneous nature of the experiment did not permit accuracy determination. The appendix A1.3 contains further details of the study and instructions.

[18]I did not find any evidence that those subjects who played the role of Player A first (and who might therefore form beliefs that Player A's entry helps Player B) had higher altruism later as Player B's than those who were in the role of Player B first. Therefore, either subjects did not think every deferral helped Player B, or they did not display positive reciprocity.

3.3.2 Evidence for Correlated Beliefs and Preferences

I observe considerable heterogeneity in entry decisions across Player A, depending on what they chose in the sub-game Player B faces (Table 15). Although this heterogeneity mirrors the pattern observed in Study 2, we cannot directly conclude that differences in beliefs drive this behavior. Since I designed all two-stage games to pay the lowest outcome to Player B if Player A decided not to defer, even if beliefs did not vary systematically across participants in the role of Player A, we would expect those players who cared more about the payoffs of others to enter more often and to pick altruistically in the role of Player B.

As Charness and Rabin (2002) also observed, in order to infer subjects' social preferences from their decisions in the trust-dictator games, we need to make assumptions about the beliefs they hold concerning Player B's behavior. Without an a priori restriction on beliefs, for example, that they equal objective probabilities, the factor that drives Player A's decisions is unclear. Is it preferences or beliefs? We can circumvent this identification problem in two ways. The first is to collect data on beliefs. The second is to collect other choice data that can help identify social preferences. Subjects' choices in the role of Player B provide such data in this study. Both sources of data help document the impact of systematic differences in beliefs on strategic actions.

Let us investigate whether the assumption that people on average share common beliefs about others can be maintained given the choice data across different games. Note that, by design, when Player A decides whether to defer or take the social allocation where he gets more than Player B, I have Player B face a sub-game where he gets less than Player A in both options. Referring back to Table 1, we see that although the preference parameters α and β inferred from choices in dictator games are positively correlated, considerable heterogeneity exists in the difference between these two parameters *within* a person. This variation, along with the aforementioned design feature, is central to the following demonstrations of the violation of the mutual consistency assumption.

First, consider a group of participants who are never willing to sacrifice to increase the other player's payoff if the other player's payoff is already higher than their own ($\beta_i = 0$). Some of these individuals, however, may be somewhat altruistic when the other's payoff is lower than their own ($\alpha_i > 0$), whereas some may still be purely selfish ($\alpha_i = 0$). Now consider these participants' entry decisions in the role of Player A in games where Player B's payoffs are higher than those of Player A's. In this decision context, Player A decides whether to take the social allocation or defer based

on his degree of altruism β_i as he is behind and his beliefs about what Player B may choose in the sub-game. If beliefs are not systematically different, the variation in α_i should not affect Player A's entry decision. However, I find that those with higher α_i's tend to enter such games more often than those who are more selfish. Similarly, subjects with higher β_i's defer to Player B in trust-dictator games where Player B is behind more often than subjects with lower β_i's do. Table 16 reports entry rates by pairs of preference parameters. The left-hand panel summarizes entry rates in games where Player A is ahead and Player B is behind, and the right-hand panel summarizes entry rates in games where Player A is behind and Player B is ahead. Since α is greater than or equal to β for almost all individuals, we do not observe every combination of these two parameters. However, a clear pattern of entry variation conditional on preferences emerges.

The results of a probit model of Player A's choice of entry in Table 17 confirm the significance of this pattern in the data. The variable "other's parameter is the subject's own preference parameter in the shoes of Player B. For example, it is α_i if Player B is ahead. The variable "own parameter" is the subject's own preference parameter as it applies to his role. For example, it is β_i if Player A is behind. The results in the first column show that, controlling for one's degree of altruism as it applies to the context he is facing, there is still a strong impact of the person's degree of altruism in the different context that Player B is facing. The second column introduces more controls. Compared with the average propensity of other similarly altruistic Player A's to enter in a gamble, those who are more altruistic in the shoes of Player B enter more often.

This finding suggests that unless this systematic difference is controlled for, entry propensities cannot be used to infer which Player A is more altruistic in the context he is facing. Consider Player A1, who is very altruistic when ahead and completely selfish when behind, and Player A2 who is moderately altruistic in both cases. Now imagine we observe their choices in trust-dictator games where they are ahead, and we are interested in their altruism in that position. As Player A1 will have a more pessimistic expectation about the outcome of the sub-game, his propensity to enter may not differ from that of Player A2, who has a more optimistic expectation. Therefore we may erroneously conclude that they have similar levels of altruism, if we assume that their beliefs are not systematically different.

Now I turn to the second model-free piece of evidence against mutually consistent beliefs. The demonstration of this evidence also relies on the variation in the within-person difference of two

preference parameters. Consider subjects who are much more altruistic when they are ahead than when they are behind (high α_i - β_i difference) versus others who do not have as high a sensitivity to relative position (low α - β difference). The subjects with a higher sensitivity to relative position in dictator games should also display a higher sensitivity in trust-dictator games. In other words, subjects with higher sensitivity to relative position display a larger disparity between their entry propensities across games where they hold different positions. If we maintain the assumption that beliefs do not systematically differ across Player As then we would expect, compared to rest of Player As, Player As with higher α - β difference to defer much more frequently in trust-dictator games where they are ahead compared to games where they are behind.

However, I find exactly the opposite. Table 18 presents the results of a binary probit regression of entry choice. The explanatory variables include subject fixed effects to capture subjects' mean propensity to defer in trust-dictator games. Question fixed effects capture the mean entry behavior across all subjects for a given trust-dictator game. I find that subjects with higher sensitivity to relative position (high α - β difference) are less likely than usual to defer when they are ahead, than are those with a lower sensitivity.

Note that we formulated our first expectation under the assumption of mutually consistent beliefs, which allowed us to make predictions about differences in choice based purely on differences in preferences. However, if Player A's beliefs about Player B's choice vary strongly based on what he would do in Player B's position then our expectations would change dramatically, as subjects with higher sensitivity to relative position (high α - β difference) would also be more sensitive to the impact of relative position on Player B's choices.

To see how we can reach the opposite prediction under the hypothesis that subjects' beliefs about Player B's choices are related to what they would have done in Player B's position, note that all the games share two important attributes by design. First, when Player A is ahead, he is ahead in all outcomes of the trust-dictator game and Player B is behind. Second, entry is more attractive if Player A is more altruistic when all else is equal. Therefore when Player A is ahead, he should have a higher incentive to enter due to his social preferences, but competing disincentive to enter due to the expectation that Player B is less likely to choose altruistically when behind. Although the disincentive may exist all subjects, the more sensitive a subject is to his own relative standing, the more he may also be sensitive to Player B's relative standing. When subjects with a

higher α - β difference put themselves in the shoes of Player B, they will predict a higher difference in Player B's propensity to choose altruistically as Player B's relative position changes. Therefore, if beliefs carry enough weight in Player A's decisions, subjects with a higher α - β difference may avoid games where they are ahead, more than those with a lower α - β difference would[19].

In appendix A3.2, I report estimates from a model recovers beliefs from observed actions. I find that the beliefs which rationalize actions correlate with preferences, replicating the evidence from previous studies that elicited beliefs in different ways.

3.3.3 Summary and Discussion of Study 3 Results

In Study 3, I provide the first strategic choice evidence congruent with the existence of heterogeneous beliefs. I show that heterogeneity in entry decisions in a strategic game is systematically related to the first mover's individual preferences in the context that the second mover faces. If the person would have been selfish in the sub-game, then he is less trusting of others. Note that people demonstrate sensitivity to the differences in the context between themselves and others. While making decisions in an advantageous position, their beliefs about the choice of the other person, who is in a disadvantageous position, are influenced by their own degree of concern for others when they are in that same disadvantageous position themselves. In other words, they are able to put themselves in the shoes of the other person and base their beliefs on what they would do under those circumstances.

The evidence from Study 3 demonstrates how significant the impact of systematic heterogeneity in beliefs can be on strategic decisions. Differences in beliefs so strongly influence differences in deferral decisions that the decisions observed are the opposite of what we would expect if we assume individuals are best responding to homogenous expectations. This impact results in strategic decisions that are seemingly inconsistent with subjects' choices in dictator games. However, systematic differences in beliefs can explain these reversals.

If we maintain the assumption of mutual consistency in beliefs is maintained for inference, two main biases arise. First, the decisions (in the role of player A) of those subjects who are moderately

[19]Even though I find no reason to expect a reversal in the sensitivity to relative standing, I also entertain the idea that social preferences in different roles might be different. Therefore, I use the elicited beliefs to capture the impact of differences in beliefs on entry behavior. When I incorporate elicited beliefs rather than maintain an assumption of mutual consistency, the relative entry behavior is consistent with the relative altruism in dictator games. This finding suggests systematic differences in beliefs, and not the impact of roles, produce the seemingly inconsistent pattern of behavior. Results are available upon request.

altruistic when ahead, but very selfish when behind will be interpreted as selfish even when ahead. I will demonstrate this bias in the following section. Second, the choices (in the role of Player A) of those subjects who have a high sensitivity to relative standing will result in the inference that they have a low sensitivity to relative standing.

4 Implications of systematic heterogeneity in beliefs

4.1 Misestimation of preferences

The findings of all three studies show that ignoring the systematic heterogeneity in beliefs about altruism is misleading. This section aims to detail the identification consequences of assuming objective probabilities for beliefs when this heterogeneity is present[20].

In the first simulation, I consider the same two-stage trust-dictator game as in Study 3. I start with a joint distribution of α and β as demonstrated on the top panel of Figure 5. For parsimony, I allow for the people who choose selfishly in the dictator sub-game to overestimate the percentage of selfish choices in the population by 10% and those who choose altruistically to underestimate it by the same amount. I simulate the choices of Player A in games where trusting increases social welfare as in Study 3 and where Player A is behind in all outcomes. I then infer Player A's β from his decisions assuming rational expectations.

The lower panel of Figure 5 displays the distribution of estimated parameters. The shading allows us to see which individuals are miscategorized and in what direction. If people hold common beliefs, then we would expect those with higher β's to defer the decision to Player B more often, as entry increases social welfare. Those with relatively higher α's, who are therefore more optimistic beliefs about Player B's altruism, are miscategorized as having higher β's.

I also repeat the exercise for games where trusting in others' altruism hurts Player B. In such games, a more altruistic Player A is less likely to enter than a more selfish Player A if they share the same beliefs. Figure 6 displays the distortion in the estimated preference distribution. The participants whose α's are small, meaning they would be selfish in the shoes of Player B, are relatively less likely to enter than if they had the correct beliefs. Under the assumption of mutual consistency, we would misinterpret this behavior as them having more concern for Player B.

[20]To highlight the role of systematic individual deviations, I do not allow for conservatism in these demonstrations.

Figure 4: Simulation depicting biases in preference estimation when trusting increases social welfare.

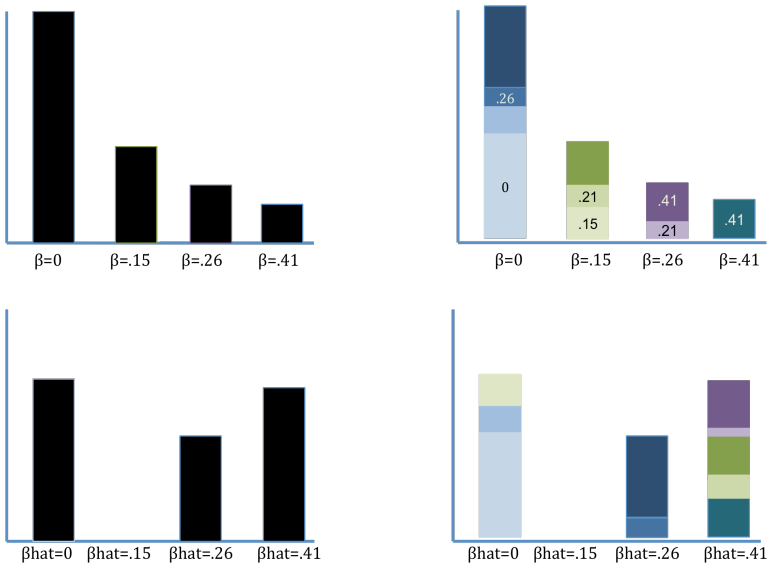
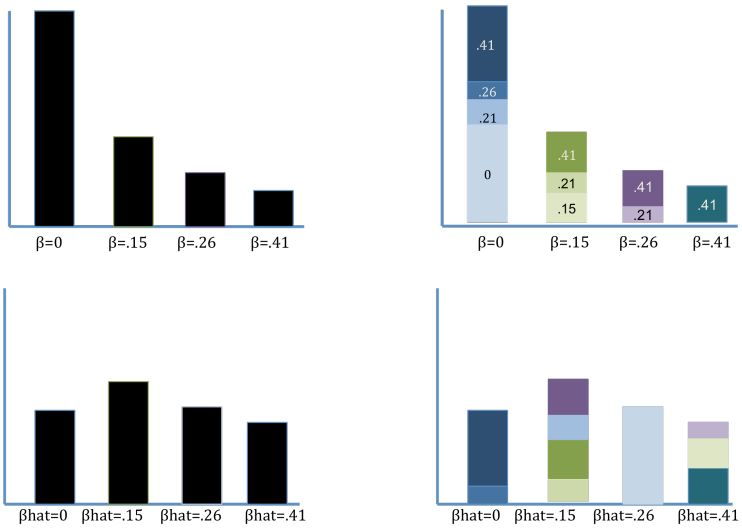


Figure 5: Simulation depicting biases in preference estimation when trusting decreases social welfare.



4.2 Broader Implications for Future Research and Discussion

The evidence of systematic heterogeneity in beliefs is congruent with the intuition that people form their belief distributions of others' choices based on how far they see themselves from choosing

each of the options. Understanding the reason for this projection is beyond the scope of this paper and does not impact its results or conclusions. We might base our educated guesses on the false consensus literature, which has provided two potential reasons for why people think more support exists for an option if they picked that option. The first is based on the premise that one's own position acts as the prior on which beliefs are based (Hoch 1987, Dawes 1989). The second is based on egocentrism and claims people are not sufficiently able to allow for the fact that even beyond informational reasons, others might be different from them (Krueger and Clement 1994). Future research can investigate whether the reason for projection impacts if and how the correlation between preferences and beliefs evolve in repeated interactions.

If projection results from informational differences, whether the differences across individuals in their beliefs are initial differences that can be eradicated by learning in repeated interactions depends on the type of information these individuals can get in equilibrium. If individuals observe others' actions, two further research questions arise given the results in this paper. First, how do people make inferences about the underlying preferences behind others' choices in order to predict their future choices more precisely? Given that peoples' ability to infer others' preferences from their observed choices suffers from the same confounds this paper discusses due to unobservable beliefs, it is important to understand their inference process. Second, do people with different preferences learn differently? Since individuals systematically differ in their beliefs, which subset of all possible sub-games they get exposed to in the market depends on these beliefs, limiting equal chance of learning from real-world realizations. Studying the role and extent of such selection in learning due to heterogeneity in initial priors can help answer the long-term market impact of the relationship between preferences and beliefs on strategic actions.

The evidence of systematic heterogeneity in beliefs may also be congruent with beliefs impacting preferences instead, due to a need to conform. In a different context, Orhun and Urminsky (2009) provide evidence that manipulated changes in preferences lead to changes in beliefs. This evidence is consistent with projection, rather than with an alternative account that beliefs influence preferences. However, the current paper does not directly test between these accounts, as the source of the correlation does not impact the contribution and implications of the paper. Regardless of its source, this correlation would impact strategic actions in the same way and inference from strategic choices would be similarly hindered. Importantly, the evidence this paper presents on the specific

form of the correlation would still be crucial for appropriately modeling it. However, if future research proves that altruism is indeed constructed from beliefs and the need to conform, then the preference specifications in the behavioral economics literature that are used to capture altruistic behavior would need to be revised to correctly specify the underlying reason for seemingly-kind behavior. This account would also raise the next important question: if differences in preferences are due to differences in beliefs, where do the differences in beliefs come from?

Related to the question of how people would form beliefs about others' preferences from their choices, we can also ask how people form beliefs about others' beliefs. When a selfish Player B observes that Player A deferred the decision to her, benefiting her outcome and risking his own, does she think Player A expected Player B to return the favor, or does she think Player A was acting mainly out of altruism? Given that beliefs about others' preferences depend on own preferences, do beliefs about beliefs differ across people, and if so, what do they relate to? Geanakoplos et al. (1989), Rabin (1993), Dufwenberg and Gneezy (2000), Gneezy et al. (2000), and Dufwenberg and Kirchsteiger (2000, 2004) noted contexts where beliefs should be incorporated directly into one's utility function, as people might care about why the other person behaved in a certain way. Extending this paper's approach to reciprocity would require models of reciprocal behavior to allow not only individual differences in social preferences, but also individual differences in beliefs about the intensions of others. Heterogeneity in beliefs about the intensions of others, therefore, will be an important concern in future research in teasing out factors that drive reciprocal behavior.

Lastly, note that what people think others' preferences are may affect how they allocate resources, even in a non-strategic setting. If welfare is defined over the utilities of the parties involved, rather than over outcomes, what a social planner or a benevolent dictator thinks others' preferences are will have an important effect on the choice of the allocation profile. As research moves toward defining welfare over utilities rather than outcomes, this issue is bound to become a centerpiece of discussion.

5 Conclusion

In a series of experiments, I document considerable heterogeneity in beliefs about others' choices in a given decision context, which is at odds with the commonly used mutual consistency assumption for making preference inferences from strategic choices. These differences in beliefs correlate with

differences in individuals' evaluations of the options in the given decision context. In a strategic game involving trust and altruism, I demonstrate that if the assumption of mutual consistency is maintained, people's strategic actions would lead to an inference of altruistic preferences that is inconsistent with preferences inferred from non-strategic actions. In particular, differences between individuals in their trust of others are the opposite of what we would expect if we assumed individuals are best responding to homogenous expectations. I show that these results are congruent with differences in expectations of altruism that result from putting one's self in the other's position. I then detail how ignoring this heteroskedasticity results in mis-estimation of the preference distribution.

In general, the solution to the problem of preference inference from strategic choices should rely on augmenting the data such that identification does not rely on assumptions about unverifiable beliefs. The first possibility is to collect additional choice data that produce variation in choices, keeping beliefs constant. For example, in the two-stage game considered in this paper, such variation can be obtained by keeping the dictator sub-game constant, while varying the payoff of Player B in the case of Player A choosing not to trust. However this type of a solution may not be generally feasible. The second solution, as Manski (2004) points out, is to measure expectations.

Especially in situations where the researchers would like to test for non-standard preferences using deviations from normative behaviors in strategic interactions, collecting data on beliefs rather than making an assumption is paramount. A priori, whether individuals are not behaving in line with the normative prescription due to sub-optimal best-responding, non-standard preferences, or due to beliefs that systematically deviate from objective beliefs is unclear. The need to separate the inability to best respond to one's beliefs from potentially wrong beliefs about others has been recognized by Stahl and Wilson (1995) and Camerer et al. (2004), but in their applications this necessity was addressed by making an alternative assumption about the particular relationship between beliefs about others and one's own ability to reason[21]. Recently, experimental studies directly elicit expectations about others' actions to address the inference problem. Expectation measurement helps Bellemare et al. (2008) address whether giving more than the normative prediction of zero in an ultimatum game is motivated by preferences or beliefs about responders' reactions. In a different context, Fischbacher and Gächter (2010) use elicited beliefs about others' contributions

[21]This assumption, motivated by the need for internal consistency, generates a correlation between the beliefs about the sophistication of others' thinking and one's own sophistication.

to separate the role of changing beliefs from the role of cooperation preferences in explaining the decline of contributions to a public good played repeatedly.

While this paper agrees with the previous literature on the necessity of eliciting beliefs, it also highlights the importance of knowing the way in which beliefs deviate from the rational expectations assumption to achieve correct inference even when belief data is available. In particular, in contexts where the correlation of beliefs and actions needs to be modeled, knowing the correct relationship between beliefs and preferences is essential in informing the requirements for data collection and statistical modeling. The results show that the relationship between preferences and beliefs is different than what previous empirical operationalizations allowed for. Differences in beliefs do not depend on differences in choices alone, as assumed in the false consensus literature, neither are they captured solely by subject-specific factors, as assumed in Bellemare et al. (2008). If researchers find it important to allow for a correlation of beliefs and actions in any given context, an individual's belief about the choice of others in a decision task should be allowed to vary with the differences between the individual's own valuations of different options in the same task. How much the individual values each option can be directly elicited or predicted based on the individual's preferences.

This paper has shown that the correlation between preferences and beliefs about others' actions will impact strategic choices and bias the estimation of preferences; thus empirical studies should appropriately take it into account. I expect the results to generalize to a broad array of uncertain decision-making contexts in which the individual has to form expectations about the behaviors of others.

References

Bellemare, Charles, Sabine Kroger and Arthur van Soest. (2008) "Measuring Inequity Aversion in a Heterogenous Population Using Experimental Decisions and Subjective Probabilities." Econometrica. Vol.76, No.4, 815-839.

Barber , Brad M., Terrance Odean (2001) "Boys Will be Boys: Gender, Overconfidence, and Common Stock Investment." Quarterly Journal of Economics. Vol. 116, No. 1, Pages 261-292.

Camerer, Colin. (1993) Behavioral Game Theory. Princeton University Press.

Camerer, Colin and Dan Lovallo (1999) "Overconfidence and Excess Entry: An Experimental Approach." The American Economic Review, Vol. 89, No. 1, pp. 306-318.

Camerer, Colin, Teck-Hua Ho and Juin-Kuan Chong (2004) "A Cognitive Hierarchy Model of Games." The Quarterly Journal of Economics, August, 861-898.

Charness, Gary and Brit Grosskopf (2001) "Relative payoffs and happiness: an experimental study" Journal of Economic Behavior and Organization Vol. 45 pages 301328.

Charness, Gary and Matthew Rabin.(2002) "Understanding Social Preferences with Simple Tests." The Quarterly Journal of Economics. August, 817-869.

Dawes, Robyn (1989) "Statistical criteria for establishing a truly false consensus effect." Journal of Experimental Social Psychology. 25:11, 1-17.

Dufwenberg, Martin and Uri Gneezy (2000) "Measuring Beliefs in an Experimental Lost Wallet Game." Games and Economic Behavior, 30, 163-182.

Dufwenberg, Martin and Georg Kirchsteiger. (2000) "Reciprocity and wage undercutting." European Economic Review. 44, 10691078.

Dufwenberg, Martin and Georg Kirchsteiger (2004) "A theory of sequential reciprocity." Games and Economic Behavior. 47 (2004) 268298.

Geanakoplos, J., Pearce, D., Stacchetti, E. (1989) Psychological games and sequential rationality. Games and Economic Behavior. 1, 6079.

Gneezy, U., Gth, W., Verboven, F., (2000). "Presents or investments? An experimental analysis." J. Econ. Psych. 21, 481493.

Fischbacher, Urs and Simon Gächter (2010). " Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments." American Economic Review, 100(1), 541-556.

Fisman, Raymond, Shachar Kariv and Daniel Markovits. (2007) "Individual Preferences for Giving" The American Economic Review, 97:5. 1858-1876.

Hoch, Stephen J. (1987) "Perceived consensus and predictive accuracy: the pros and cons of projection" Journal of Personality and Social Psychology, vol. 53, no2, pp. 221-234.

Huck, Steffen and Georg Weizsacker (2000). "Do players correctly estimate what others do? Evidence from conservatism of beliefs." Journal of Economics Behavior and Organization. Vol.47, 71-85.

Iriberri, Nagore and Pedro Rey-Biel. (2009) "Elicited Beliefs and Social Information in Modified Dictator Games: What Do Dictators Believe Other Dictators Do?", Working Paper, Universitat Pompeu Fabra.

Krueger, Joachim and Russell W. Clement (1994). "The truly false consensus effect: An ineradicable and egocentric bias in social perception." Journal of Personality and Social Psychology, Vol 67, issue 4, p. 596-61.

Malmendier, Ulrike and Geoffrey Tate (2008) "Who makes acquisitions? CEO overconfidence and the market's reaction" Journal of Financial Economics, Volume 89, Issue 1, Pages 20-43.

Manski, Charles. (2002) "Identification of decision rules in experiments on simple games of proposal and response." European Economic Review. Vol 46, 4-5, 880-891.

Manski, Charles. (2004) "Measuring Expectations." Econometrica, Vol. 72, No. 5, 1329-1376.

Marks, Gary and Norman Miller (1987). "Ten years of research on the false consensus effect: An empirical and theoretical review". Psychological Bulletin, vol 102, 729.

Orhun, Yesim and Oleg Urminsky (2009) "How Self Impacts Beliefs about Others: The Role of Assumed Correspondence of Evaluations Underlying Choice". Working Paper, The University of Chicago.

Rabin, Matthew. (1993) "Incorporating Fairness into Game Theory and Economics." The American Economic Review. Vol. 83, No. 5, 1281-1303.

Stahl, Dale and Paul Wilson (1995) "On Players' Model of Other Players: Theory and Experimental Evidence." Games and Economic Behavior 10, 218-254.

Selten, Reinhard and Axel Ockenfels (1998) "An experimental solidarity game" Journal of Economic Behavior and Organization Vol. 34, pages 517-539.

Table 1. Joint frequency distribution of degrees of altruism when the decision maker is ahead or behind in relative standing to the recipient.

	beta (behind)							
alpha (ahead)	0	0.12	0.15	0.21	0.26	0.34	0.41	Total
0	23	0	0	0	0	0	0	23
0.12	3	0	0	0	0	0	0	3
0.15	3	3	0	1	0	0	0	7
0.21	2	0	2	1	0	1	1	7
0.26	2	4	2	0	0	1	0	9
0.34	7	0	1	2	3	0	0	13
0.41	13	1	6	2	2	4	4	32
Total	53	8	11	6	5	6	5	94

Table 2. Study 1, Session 1. Actual Play, Average Beliefs, and Beliefs by Own Action in the given game. (200 Tokens=$1)

Selfish Option		Sacrifice Option		%	Average	Avg. Beliefs	Avg. Beliefs
Self	Other	Self	Other	Selfish	Beliefs	were selfish	sacrificed
850	200	750	600	0.33	0.58	0.78	0.48
650	200	600	300	0.33	0.56	0.71	0.49
750	150	700	450	0.17	0.40	0.58	0.36
900	150	800	450	0.29	0.57	0.70	0.52
700	300	600	600	0.42	0.54	0.64	0.47
500	150	450	450	0.25	0.54	0.78	0.46
750	0	650	200	0.33	0.49	0.64	0.42

Table 3. Study 1, Session 2: Actual Play, Average Beliefs, and Beliefs by Actual or Potential Action in the given game. (200 tokens= $1)

Selfish Option		Sacrifice		%	Average	Avg. Beliefs	Avg.	Avg. Beliefs	Avg. Beliefs
Self	Other	Self	Other	Selfish	Beliefs	if sacrificed	if selfish	if would sacrif	if would
750	200	650	400	0.58	0.56			0.51	0.62
900	0	800	300	0.21	0.40	0.35	0.61	0.24	0.57
800	300	700	600	0.46	0.54	0.47	0.75	0.34	0.61
500	150	450	450	0.38	0.47	0.37	0.65		
700	200	550	400	0.75	0.64			0.48	0.74
650	200	600	300	0.33	0.52			0.48	0.57
750	0	700	100	0.29	0.29	0.21	0.51		
850	200	750	600	0.38	0.42	0.31	0.59		
700	200	600	600	0.21	0.41	0.32	0.56	0.34	0.63
900	200	850	400	0.38	0.41			0.37	0.46
650	200	600	500	0.17	0.48			0.39	0.83
650	300	600	600	0.29	0.31	0.21	0.57		
900	0	600	400	0.54	0.48	0.26	0.65		
500	200	400	400	0.38	0.58	0.49	0.72		
750	0	700	300	0.15	0.30	0.26	0.57		
950	0	750	300	0.40	0.53	0.44	0.72		

Table 4. Study 1, Session 3: Actual Play, Average Beliefs, and Beliefs by Potential Action in the given game. (200 Tokens=$1)

Selfish Option		Sacrifice Option		%	Average	Avg. Beliefs	Avg. Beliefs
Self	Other	Self	Other	Selfish	Beliefs	if would sacrif	if would selfish
750	350	700	500	0.27	0.57	0.53	0.61
300	350	250	500	0.47	0.72	0.57	0.77
600	200	500	400	0.47	0.74	0.63	0.83
350	400	300	800	0.27	0.60	0.55	0.69
400	500	200	800	0.87	0.77	0.40	0.80
150	200	50	600	0.80	0.70	0.68	0.71
750	300	700	600	0.13	0.52	0.45	0.68
600	650	550	750	0.80	0.74	0.65	0.78
750	150	700	550	0.27	0.59	0.53	0.82
750	200	650	500	0.53	0.55	0.49	0.68
700	300	600	500	0.47	0.52	0.30	0.71
800	200	700	600	0.27	0.50	0.44	0.68
250	300	200	600	0.47	0.48	0.35	0.55
850	250	750	400	0.40	0.58	0.25	0.70
250	300	200	700	0.40	0.54	0.46	0.77
250	300	150	500	0.73	0.68	0.56	0.71
200	200	100	500	0.67	0.74	0.55	0.84
850	300	800	700	0.27	0.37	0.31	0.73
300	450	200	600	0.93	0.75	0.20	0.79
800	300	750	600	0.27	0.42	0.37	0.77

Table 5. Study 1. Tobit regression of elicited beliefs about propensity of selfishness.

VARIABLES	(1)	(2)	(3)	(4)
rho (choice=selfish)	0.284***	0.089**	0.114***	0.053*
	(0.024)	(0.039)	(0.027)	(0.031)
delta (preference intensity)		0.678***		0.493***
		(0.112)		(0.116)
Constant	0.518***	0.617***	0.827***	0.877***
	(0.054)	(0.055)	(0.093)	(0.092)
Observations	521	521	521	521
Question FE's	yes	yes	yes	yes
Subject FE's			yes	yes
Log Likelihood	-56.46	-38.66	127.5	134.8

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6. Table 4, Column 1, broken down over the sessions.

VARIABLES	Session 1	Session 2	Session 3
rho	0.314***	0.345***	0.280***
	(0.061)	(0.048)	(0.028)
Constant	0.429***	0.631***	0.460***
	(0.033)	(0.031)	(0.020)
Observations	84	148	289
Question FE's	yes	yes	yes
Log Likelihood	-11.64	-44.87	-19.01

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 7. Study 1. Session 3. Tobit regression of elicited beliefs about propensity of selfishness. Differences in within-person belief differences regarding concern withdrawal.

VARIABLES	(1)	(2)
$(\alpha - \beta)$*(dictator ahead)	-0.853***	-0.743***
	(0.224)	(0.145)
dictator ahead		-0.066*
		(0.038)
$(\alpha - \beta)$		0.138
		(0.145)
Constant	0.696***	0.671***
	(0.012)	(0.027)
Log Likelihood	35.26	-52.78
Observations	298	298
Question FE's	yes	no
Subject FE's	yes	no

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 8. Study 1. Session 2. Impact of having taken an action in the task about which beliefs are elicited.

VARIABLES	(1)	(2)	(3)
rho	0.318***	0.053	0.045
	(0.069)	(0.092)	(0.107)
delta		0.802***	0.753***
		(0.222)	(0.255)
nu 1 ((new=0) * (choice=selfish))	0.069		-0.082
	(0.108)		(0.225)
nu 2 ((new=0) * preference intensity)			-0.522
			(0.566)
nu 3 ((new=0))	0.131		0.094
	(0.104)		(0.121)
Constant			
Log Likelihood	-39.372	-38.590	-37.338
Observations	148	148	148
Question FE's	yes	yes	yes

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 9. Study 1. Differences in the correlation between preferences and beliefs based on person's error-proneness in his choices in modified dictator games.

VARIABLES	(1)
rho	0.3822***
	(0.035)
omega * (choice=selfish)	-0.782***
	(0.198)
omega	0.184
	(0.128)
Constant	
Log Likelihood	-48.029
Observations	521
Question FE's	yes

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 10. Study 2 descriptive statistics.

Subgame (Selfish)		Subgame (Sacrifice)		%	sure	certainty equivalent / entry % at the sure payoff		
Self $	Other $	Self $	Other $	selfish	payoff	avg	if selfish	if sacrificed
8.50	3.00	8.00	7.00	11.5				
2.50	3.00	2.00	7.00	61.5	$4.50	4.29 / 42%	4.16 / 37%	4.44 / 48%
6.50	2.00	6.00	5.00	15.4	$3.00	3.71 / 82 %	3.50 / 70 %	3.80 / 85 %
2.50	3.00	2.00	6.00	58.7	$4.25	4.00 / 32 %	3.74 / 19%	4.58 / 46%
7.50	3.50	7.00	5.00	26.9				
3.00	3.50	2.50	5.00	80.8	$4.25	3.86 /18%	3.85 / 15%	3.88 / 30%
5.00	2.00	4.00	4.00	26.9	$3.00	2.80 /30%	2.63 / 17%	2.88 / 38%
6.00	6.50	5.50	7.50	84.6				
7.50	1.50	7.00	5.50	20.0	$3.00	3.55 /76%	2.65 / 50%	3.72 / 81 %
3.50	4.00	3.00	8.00	48.0				
8.00	3.00	7.50	6.00	24.0				
2.50	3.00	2.00	6.00	48.0				
8.00	3.00	7.00	6.00	36.0	$4.25	4.19 / 50%	4.01 / 38 %	4.27 / 56%
2.00	2.00	1.00	5.00	80.0				
7.00	3.00	6.00	5.00	48.0				
2.50	3.00	1.50	5.00	84.0	$4.00	3.44 / 4%	3.34 / 0%	3.45 / 4.7%

Table 11. Study 2. Probit model of entry in gambles with known probability.

VARIABLES	(1)	(2)	(3)
altruism (ahead)	-0.039		-0.002
	(0.075)		(0.087)
altruism (behind)		-0.073	-0.072
		(0.076)	(0.088)
Constant			
Log Likelihood	-107.75	-107.42	-107.42
Observations	204	204	204
Gamble FE's	yes	yes	yes

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 12. Study 2. Ordered probit model of certainty equivalents in titrations of gambles with known probability.

VARIABLES	(1)	(2)	(3)
altruism (ahead)	-0.046		-0.067
	(0.056)		(0.064)
altruism (behind)		0.008	0.042
		(0.055)	(0.063)
Constant			
Log Likelihood	-398.94	-399.27	-398.72
Observations	203	203	203
Gamble FE's	yes	yes	yes

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 13. Study 2. Probit model of entry in gambles on altruism.

VARIABLES	(1)	(2)	(3)
own choice = selfish	-0.517***	-0.134*	-0.556**
	(0.149)	(0.065)	(0.238)
preference intensity		-0.974*	
		(0.433)	
Constant	-0.338	-0.369	-0.376
	(0.260)	(0.199)	(0.587)
Gamble FE's	yes	yes	yes
Subject FE's			yes
Log Likelihood	-211.45	-189.33	-139.22
Observations	408	408	360

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 14. Study 2. Probit model of entry in gambles on altruism, for each condition.

	Experiment A first		Experiment B first	
VARIABLES	(1)	(2)	(3)	(4)
own action = selfish	-0.489**	-0.714**	-0.572***	-0.440
	(0.213)	(0.296)	(0.221)	(0.306)
new * own action = selfish		0.442		-0.296
		(0.393)		(0.453)
new		-0.461		-0.077
		(0.238)		(0.282)
Constant	-0.288	-0.002	-0.391	-0.420
	(0.261)	(0.317)	(0.300)	(0.350)
Gamble FE's	yes	yes	yes	yes
Log Likelihood	-112.50	-111.15	-94.36	-93.78
Observations	224	224	161	161

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 15. Study 3 descriptive statistics.

Subgame (Selfish)		Subgame (Sacrifice)		Stay-Out Payoffs		% selfish		% enter	
Self	Other	Self	Other	Self	Other		avg	if selfish	if sacrificed
750	350	700	500	375	600	28.13	65.63	44.4	73.9
850	300	800	700	400	700	31.25	75.00		
750	150	700	550	300	600	31.25	56.25	30.0	68.1
750	300	700	600	400	600	31.25	65.63		
800	200	700	600	350	600	33.42	54.69	29.17	67.27
800	300	750	600	350	650	34.38	71.88	45.5	85.7
750	350	700	500	375	600	40.63	81.25		
700	300	600	500	350	500	40.63	62.50	53.8	68.4
800	200	700	600	300	600	42.19	59.38	41.48	75.45
750	200	650	500	300	550	43.75	40.63		
750	200	650	500	300	550	43.75	56.25	35.7	72.2
850	250	750	400	275	650	50.00	43.75		
350	400	300	800	550	200	56.25	21.88		
600	200	500	400	275	400	56.25	59.38		
250	300	200	700	400	150	59.38	43.75		
350	400	300	800	550	200	59.38	37.50	15.8	69.2
250	300	200	600	400	150	64.06	21.88	10.15	40.77
250	300	200	600	350	150	67.19	51.56	32.47	90.00
250	300	200	700	400	150	68.75	15.63	9.1	30.0
800	150	600	450	200	500	71.88	46.88	39.1	66.7
300	350	250	500	375	200	78.13	37.50		
400	500	300	900	650	200	81.25	25.00		
150	200	50	600	300	50	81.25	25.00	11.5	83.3
600	650	550	750	700	500	84.38	25.00		
300	450	200	600	475	100	87.50	50.00		
200	200	100	500	300	50	90.63	15.63	13.8	33.3
250	300	150	500	350	100	96.88	12.50	12.9	0.0
400	500	200	800	550	150	96.88	31.25	29.0	100.0

Table 16. Study 3. Player A entry rates by pairs of preference parameters for two types of games (different relative position of players).

Player A ahead, Player B behind			Player A behind, Player B ahead		
alpha	beta	entry rate	beta	alpha	entry rate
0	0	0.146	0	0	0.382
0.12	0	0.375	0	0.12	0.438
0.15	0	0.208	0	0.15	0.5
0.15	0.12	0	0	0.21	0.5
0.15	0.21	0.375	0	0.26	0.688
0.21	0	0.188	0	0.34	0.55
0.21	0.15	0.75	0	0.41	0.693
0.26	0	0.188	0.12	0.15	0.5
0.26	0.12	0.125	0.12	0.26	1
0.26	0.15	0.5	0.15	0.21	0.75
0.34	0	0.275	0.15	0.26	0.75
0.34	0.15	0.25	0.15	0.34	0.375
0.34	0.21	0.625	0.15	0.41	0.833
0.34	0.26	0.75	0.21	0.15	0.625
0.41	0	0.193	0.21	0.34	0.625
0.41	0.15	0.417	0.21	0.41	0.875
0.41	0.21	0.688	0.26	0.34	0.688
0.41	0.26	0.25	0.26	0.41	1
0.41	0.34	0.875	0.34	0.41	1
0.41	0.41	0.75	0.41	0.41	0.833

Table 17. Study 3. Probit model of Player A's entry decisions. Impact of altruism in the shoes of Player B.

Variables	(1)	(2)
	enter	enter
other's parameter	3.797***	3.142***
	(0.277)	(0.375)
own parameter FE	yes	yes
question FE		yes
question*own param FE		yes
Constant	-0.777***	-1.099*
	(0.077)	(0.586)
Observations	944	760
Log Likelihood	-544	-406.1

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 18. Study 3. Probit model of Player A's entry decisions. Within-person difference in entry propensity as a function of sensitivity to Player B's relative position in the subgame.

Variables	(1)
	enter
(α - β)*(Player A ahead, Player B behind)	-2.174***
	(0.700)
subject FE	yes
question FE	yes
Constant	-0.607
	(0.375)
Observations	800
Log Likelihood	-387.1

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1