Pre-Analysis Plan: Norms in Dictator Games and the Prisoners' Dilemma

Joakim Semb*

March 21, 2022

1 Introduction

This document presents the planned analyses for an experiment in which norms are elicited for two variations each of two commonly studied games: dictator games and the prisoners' dilemma (PD). The experiment addresses open questions in each game.

2 The PD Disjunction Effect

Shafir and Tversky (1992) showed that subjects playing prisoners' dilemmas are substantially more likely to cooperate when they do not know their opponent's choice of action (37%), than when they know that the opponent has cooperated (16%) or defected (3%). They refer to this ostensible violation of the sure-thing principle as the the Disjunction Effect. A recent replication attempt by Semb (2022) found that players under uncertainty cooperated to the same extent as players facing certain cooperation, while those facing certain defection cooperated much less; a finding

*Stockholm School of Economics; e-mail: joakim.semb@phdstudent.hhs.se

which cannot be explained by the belief data. This replication reaffirms the finding that standard preferences and beliefs cannot explain cooperation rates across game variants.

This experiment elicits norms over actions (Cooperate, Defect) in the simultaneous-move as well as sequential PD. If the social appropriateness of the two actions is not different between the situations, then this would explain the results of Semb (2022). If in fact cooperation is *more* appropriate relative to defection under uncertainty compared to when facing certain cooperation, then even the findings of Shafir and Tversky (1992) are explained.

3 Role Uncertainty in the Dictator Game

Iriberri and Rey-Biel (2011) show that in discrete-decision dictator games, prosocial behavior is greatly increased by the presence of role uncertainty, i.e. when both players make dictator decisions and it is randomly determined whose decision is actually implemented.

Semb (2022) finds no evidence that these effects stem from any form of magical reasoning; that is, subjects do not give more because they anticipate that they will then receive more. This experiment investigates whether the effect stems from preferences rather than beliefs, due to a difference in norms across the role structure variants. It also investigates the role of self-serving biases by eliciting norms from players as well as observers.

4 Sample and design

I will recruit 1000 participants using the recruitment platform Prolific. The selection criteria were a 99% (or higher) approval rate as well as current residence in the US.

Half of the subjects will be players. I will study two games: the PD as well as binary dictator games. In the latter, decisionmakers face a series of 10 binary dictator decisions, with the decisionmaker choosing one of two token allocations

between themselves and the receiver. In each decision, there is a "selfish" option and an "unselfish" option, with the latter implying a lower payoff for the decider but a higher payoff for the receiver. The unselfish option is always the efficient one.

I experimentally vary whether participants face a simultaneous-move PD, or a sequential-move PD in which the first-mover has cooperated. For the dictator games, I vary whether decisions are made under role certainty or role uncertainty. I randomly vary whether participants see the PD first, or the dictator game. Within the dictator game, the order of the decision screens is also random.

The experiment will use the Krupka and Weber (2013) procedure to elicit norms over actions in the games. Participants are asked to indicate the social appropriateness of each action on a four-point scale (Very Socially Inappropriate, Somewhat Socially Inappropriate, Somewhat Socially Appropriate, Very Socially Appropriate) and are paid for matching the most common response among participants.[1] The responses are numerically coded to $(-1, -1/3, 1/3, 1)$ respectively and denote $\nu_{A,i}$ as the numerical rating given by participant i to action A. From this we obtain mean appropriateness score μ_A for each action A. Since the experiment uses only binary actions, we can define the relative appropriateness score $\mu_A - \mu_B$ for each action pair.

Observers perform only the appropriateness rating task. Players first play both games, and then perform the rating task for both games. Players are paid either for one of the games, or for the rating task.

4.1 Main statistical tests

4.1.1 Prisoners' Dilemma

T_1. I test whether the relative social appropriateness of cooperation $\mu_C - \mu_D$ is different under strategic certainty (sequential PD facing cooperation) than under strategic uncertainty (simultaneous-move PD). I test this using a two-sided t-test of means on $\mu_C - \mu_D$ between the two treatment conditions.

T_{1b}. As a secondary point of interest, I test whether the social appropriateness

[1] One action in one situation is randomly chosen for payment.

scores of cooperation is different between the treatment conditions. Two-sided t-tests.

T_{1c}. As T_{1b} but for defection.

These tests are carried out separately for players and observers.

4.1.2 Dictator Game

For the dictator games, I define the social appropriateness of giving, as rated by participant i, as the average of $\nu_{G,i}$ across the 10 decisions. μ_G is defined as the average of these averages.

T_3. I hypothesize that giving is more socially appropriate, compared to keeping, under role uncertainty. I test this separately for observers and players using one-sided t-tests of means, on the difference $\mu_G - \mu_K$.

T_{3b}. I hypothesize that giving is more socially appropriate under role uncertainty. I test this using a one-sided t-test on μ_G.

T_{3c}. As T_{3b} but for keeping.

T_4. I hypothesize that due to self-serving bias among players, observers perceive giving to be more socially appropriate compared to keeping, than do players. As T_3, but comparing means across the player/observer dimension rather than role certainty / role uncertainty.

T_5. I hypothesize that self-serving bias is stronger under role certainty; that is, that the difference in perceived norms across the player-observer dimension is greater under role certainty. Or, equivalently, that the difference in perceived norms across the role dimension is greater among players than among observers I test this by estimating the equation

$$\nu_{G,i} - \nu_{K,i} = \alpha + \beta_1 obs_i + \beta_2 RU_i + \beta_3 (obs_i \times RU_i) + \epsilon_i$$

where obs_i is a dummy variable equal to 1 if participant i is an observer rather than a player, and RU_i is a dummy variable equal to 1 for participants under role uncertainty. The hypothesis is that $\beta_3 > 0$. One-sided test.

References

Iriberri, Nagore and Pedro Rey-Biel, "The role of role uncertainty in modified dictator games," *Experimental Economics*, 2011, *14* (2), 160–180.

Krupka, Erin L and Roberto A Weber, "Identifying social norms using coordination games: Why does dictator game sharing vary?," *Journal of the European Economic Association*, 2013, *11* (3), 495–524.

Semb, Joakim, "Magical Thinking and Generous Behavior," *Working Paper*, 2022.

Shafir, Eldar and Amos Tversky, "Thinking through uncertainty: Nonconsequential reasoning and choice," *Cognitive Psychology*, 1992, *24* (4), 449 – 474.