Ridinger, Garret

Article

Intentions versus outcomes: Cooperation and fairness in a sequential prisoner's dilemma with nature

Games

Provided in Cooperation with:
MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Ridinger, Garret (2021) : Intentions versus outcomes: Cooperation and fairness in a sequential prisoner's dilemma with nature, Games, ISSN 2073-4336, MDPI, Basel, Vol. 12, Iss. 3, pp. 1-30,
https://doi.org/10.3390/g12030058

This Version is available at:
https://hdl.handle.net/10419/257540

games

MDPI

Article

Intentions versus Outcomes: Cooperation and Fairness in a Sequential Prisoner's Dilemma with Nature

Garret Ridinger

Department of Management, College of Business, University of Nevada, Reno, 1664 N Virginia St., Reno, NV 89557, USA; gridinger@unr.edu

Abstract: This paper investigates the importance of concerns about intentions and outcomes in a sequential prisoner's dilemma game with nature. In the game, there is a chance that the first mover's choice is reversed. This allows the separation of intended actions from the resulting outcomes. Equilibrium predictions from theoretical models of fairness are tested experimentally by varying the chance the first mover's choice is reversed and whether the second mover observes the first mover's choice. The results show that second mover cooperation is higher when the first mover has little control over their choice and when the second mover is not told what the first mover chose. While subject behavior is consistent with concerns for both intentions and outcomes, the results indicate that these concerns work in ways not predicted by current theoretical models. In addition, I find that psychometric measures of empathic concern and perspective taking are correlated with second mover cooperation and provide potential explanations for the experimental results.

Keywords: fairness; intentions; cooperation; reciprocity; empathy; prisoner's dilemma game

Citation: Ridinger, G. Intentions versus Outcomes: Cooperation and Fairness in a Sequential Prisoner's Dilemma with Nature. *Games* **2021**, *12*, 58. https://doi.org/10.3390/g12030058

Academic Editors: Ulrich Berger and Cristina Bicchieri

Received: 13 June 2021
Accepted: 12 July 2021
Published: 22 July 2021

1. Introduction

Experimental evidence indicates that people often deviate from maximizing their own monetary payoff. In the one-shot sequential prisoner's dilemma, if each player maximizes her own payoff, then the equilibrium prediction is that both players defect. Despite this, cooperation rates by both players are significant [1–3]. To explain cooperation in one-shot games, researchers have suggested that people care about fairness and have incorporated these concerns into game-theoretic models [4–12]. These theoretical models of fairness can be separated into two types: outcome based and intention based.[1] Outcome-based models capture concerns over distributions. An example of an outcome-based model is inequity aversion [6] which allows people to compare their payoff with others and prefer payoffs that are more equal. Intention-based models allow beliefs about others actions to influence fairness concerns. An example of an intention-based model is the reciprocity model [8], which captures that people may prefer to be kind to people who are kind to them and punish people who are unkind to them. Both modeling approaches incorporate fairness concerns that are likely important in a wide range of human behavior, but in certain environments the predictions by the two approaches can be quite different. The distinction between intentions and outcomes has real-world applications. For example, in the United States legal code, there are different consequences for being charged with involuntary manslaughter compared to first-degree murder. While the outcome is the same in both cases, the intention behind the homicide matters. Despite it's importance, it is still not fully understood how the the relative strength of individual concerns about intentions and outcomes influences human behavior.

Conditional cooperation in the sequential prisoner's dilemma is consistent with concerns for outcomes [6] and intentions [8]. Due to this, prior research has been unable to disentangle the two effects to understand their importance in explaining cooperation [1–3]. This paper introduces a novel game called the sequential prisoner's dilemma with nature. In the game,

the first mover decides whether to cooperate or defect. After the first mover's choice, there is a chance the choice is reversed by nature. After observing both what the first mover chose and the results from nature, the second mover can choose to cooperate or defect. This creates a situation where the first mover may intend to cooperate but due to chance they end up defecting. Since the second mover observes what the first mover intended to do and the outcomes are kept the same, the game can differentiate between the two fairness approaches. The game captures environments in which there is an imperfect correlation between actions and the results of those actions. Therefore, it can shed light on a wide range of situations including principle agent problems. For example, an employee can choose to work hard (cooperate) or not work hard (defect) on a project. Hard work does not guarantee that the project will be profitable for the employer but it could make it more likely that the project is a success. After observing the effort level and whether the project was successful, the employer may choose whether to reward the employee (cooperate) or not reward the employee (defect).

Often the intentions of others are not fully observable. To explore how information influences cooperation, this paper introduces a variant of the sequential prisoner's dilemma with nature where the second mover is not told what the first mover chose, but does know the results of nature. This feature captures situations where the second mover must infer the first mover's intended choice based on the first mover's control over their choice and the results of nature. For example, in principal agent problems the employee effort level is often not observable. Instead, the employer only observes whether the project was successful. The employer must infer the effort level that the employee contributed based on the correlation between effort and the success of the project. Comparing the two games can add to our understanding of how information about the person's intended choice influences individual behavior.

This paper examines the relative influence of intentions and outcomes on cooperation. Specifically, I ask: what do existing fairness models predict as information and control changes in the sequential prisoner's dilemma with nature? How well do the models capture what people actually do? To address these questions, I begin theoretically by analyzing the equilibrium predictions of outcome-based, intention-based, and combined fairness models. The modeling approaches predict different equilibrium behavior depending on individual types of players and their preferences. I test the theoretical predictions empirically using a laboratory experiment. The design of the experiment allows the separation of intentions and outcomes as well as tests the role of information.

Theoretical results under perfect information show that the outcome-based model of inequity aversion [6] predicts that second mover cooperation depends only on the results of nature. Suggesting that cooperation will be unaffected by the first mover's choice. The intention-based model of reciprocity [8] predicts that cooperation depends only on the first mover's choice and that the results of nature will not affect equilibrium behavior. Specifically, inequity aversion suggests that changes in information or control will not influence equilibrium behavior. However, reciprocity predicts that second mover cooperation will be higher when control by the first mover increases, and cooperation should increase when there is imperfect information about the first mover's choice. To account for the possibility that individuals may care about both outcomes and intentions, I introduce the mixed-concerns model. Using psychological game theory, the mixed-concerns model combines both inequity aversion [6] and reciprocity [8] into a single framework. The model allows for heterogeneity in subject's weight of two concerns. If individuals care about both reciprocity and inequity aversion, then there exists an additional equilibrium depending on the relative strength of the two concerns.

These predictions are tested experimentally by varying the chance the first mover's choice is reversed and whether the second mover observes the first mover's choice. The results show that second mover cooperation is higher when the first mover has little control over their choice and when the second mover is not told what the first mover chose. While subject behavior is consistent with concerns for both intentions and outcomes, the results

indicate that these concerns work in ways not predicted by current theoretical models. Specifically, conditional cooperation by second movers was higher when control was low. This result is puzzling as it is opposite of what is predicted by models of reciprocity. Using psychometric measures, I find that differences in perspective taking ability provide a potential explanation for the puzzle. In addition, higher empathic concern is found to be correlated with increased conditional cooperation by second movers.

Previous research on the sequential prisoner's dilemma has found that second movers are more likely to cooperate if the first mover cooperates [1–3]. This finding is in line with other evidence that conditional cooperation is an important explanation for behavior in social dilemmas [13–17]. Using the sequential prisoner's dilemma, Dhaene and Bouckaert [18] find evidence that conditional cooperation by second movers matches the theoretical predictions from the reciprocity model of [8] while Blanco et al. [19] show that individual measures of inequity aversion [6] can predict second mover behavior. As a result, it is still unclear whether second mover conditional cooperation in the sequential prisoner's dilemma is due to intention-based reciprocity or outcome-based concerns.

In many game-theoretic situations, intention-based models and outcome-based models give similar predictions. This can make it difficult to examine which concerns may have lead to the observed experimental behavior. One approach to has been to vary alternative choices players could have chosen. Results from Falk et al. [20], Bolton and Okenfels [21], and Falk and Kosfeld [22] suggest that changes in the the alternatives available may have influenced behavior. However, changes in the alternatives seemed to have little or no effect in Stanca [23] and Charness and Rabin [10]. Another approach compares a treatment where a subject has full control over their choice to a treatment where the subject has no control over their choice. Typically, the subject's choice in the no control treatment is selected via random device. Using this approach, Charness [24] and Falk et al. [25] found that intentions were important in explaining subject behavior, while Bolton et al. [26] found that only outcomes mattered. One potential issue in these experiments is that if individuals feel fundamentally different towards random devices compared to when a person is making a choice, then there could be a confounding variable that may bias the results. To control for this, I keep the random device in all the treatments. What varies is the chance the first mover's choice is reversed. In addition to controlling for potential bias, this feature creates a more realistic situation where people have more or less control, but their intended choices still matter.

The experiment conducted by Charness and Levine [27] used both a random device and varied the alternatives available. Using a modified gift exchange game, the experiment included a coin flip that determined whether the wage of the employee would be higher or lower than what the employer chose. The potential payoffs for the employee were the same, whether the employer chose a high wage and chance made it lower or the employer chose a low wage and chance made it higher. The results suggest that the intentions of the employer influenced the wage choices by employees. Charness and Levine [27] did not keep all the potential outcomes constant. Instead the alternatives that could be reached were either very beneficial for the employer or very beneficial for the employee. This is how an employer's choice was viewed as having good intentions or not. One key difference in this paper is that in the experiment the potential end node payoffs are kept constant irrespective of the first movers choice. The first mover can only make certain outcomes more likely to occur.

Information about what the first mover chose can be potentially important in understanding fairness. Charness and Levine [27] did not include a treatment where workers did not know what wage the firm selected. While this is realistic in the context of their experiment, generalizing the results to other domains becomes difficult in situations were the first mover's choice is not observable. For example, using a trust game Cox and Deck [28] examined second mover behavior when the first mover's choice had a 25% chance to be reversed, but the second mover was not told what the first mover chose. The results suggest that second movers gave the first movers the benefit of the doubt. Cox and Deck [28] did

not include a treatment where the first mover's choice was known to the second mover. In this paper, I include a condition where the second mover is told what the first mover chose and a condition where the second mover is not told that information. Potential changes in behavior between these two treatments can shed light on the importance of observing others' intentions.

2. Sequential Prisoner's Dilemma with Nature

Figure 1 shows the sequential prisoner's dilemma with nature under perfect information. The first mover decides whether to cooperate or defect first. After first mover chooses, nature will randomly select cooperate or defect. After observing both what first mover chose and the choice by nature, the second mover can choose to cooperate or defect. Figure 2 shows the game with imperfect information. In this game, the choice by first mover is no longer observed by the second mover. The second mover only observes whether nature has cooperated or defected.

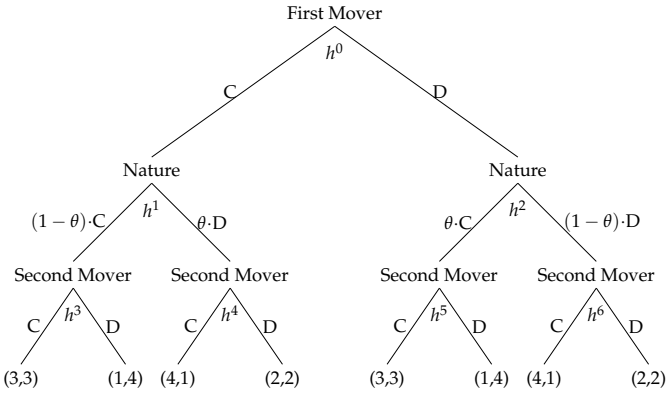

Figure 1. Sequential Prisoner's Dilemma with Nature and Perfect Information.

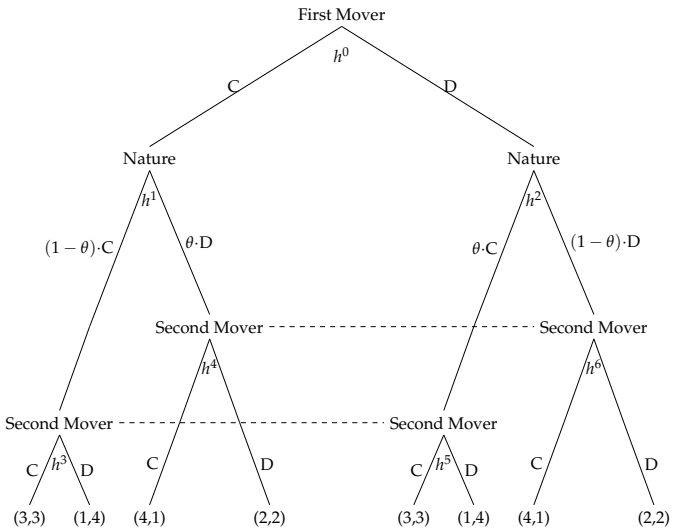

Figure 2. Sequential Prisoner's Dilemma with Nature and Imperfect Information.

The probabilities that nature will choose cooperate or defect differ depending on the choice of player 1. The term θ is the chance that the first mover's choice is reversed.[2] When $\theta < \frac{1}{2}$ and the first mover cooperates, there is a higher chance that nature will cooperate compared to if the first mover choose to defect. Natures choice can be thought of as the first mover's control. Lower values of θ make it more likely that that nature will cooperate if the first mover cooperates and defect if the first mover defects. This paper will focus on the case where $\theta \leq \frac{1}{2}$.[3]

3. Theories of Social Preferences

This section provides a review of the theoretical predictions of outcome-based and intention-based models of fairness. For interested readers, formal definitions and equilibrium predictions for the discussed models can be found in the Appendix A. The analysis focuses on second mover behavior, but first mover equilibrium predictions are available upon request.

Proposition 1. *Inequity aversion [6] predicts that second mover cooperation only depends on the results of nature and information about the first mover's choice will have no effect on cooperation rates.*

A prominent model of outcome based preferences is the model of inequity aversion [6]. In a two player game where players have inequity averse preferences [6], each individual i has the following utility function:

$$U_i(\pi_i, \pi_j) = \pi_i - \alpha_i \cdot \max\{\pi_j - \pi_i, 0\} - \beta_i \cdot \max\{\pi_i - \pi_j, 0\} \quad (1)$$

where $\alpha_i, \beta_i \geq 0$ and the payoff for individual i is π_i and the payoff for individual j is π_j. Both α_i and β_i capture the degree to which individuals dislike inequality that is advantageous and disadvantageous, respectively. The model captures the idea that people prefer distributions that are more equal.

With perfect information, inequity aversion predicts that cooperation by second movers will only occur if nature cooperates (see Proposition A1 in Appendix A for details). In addition, second mover cooperation should not differ depending on whether the first mover cooperated or defected. For second movers, the key parameter that can lead to cooperation is β_i. In order for the second mover to cooperate they must sufficiently dislike getting more than the first mover.

Under imperfect information, inequity aversion predictions are the same. This occurs because the second mover is only concerned about the distribution of outcomes at the end node of the game. As a result, the decision to cooperate is based entirely on the end node payoffs and not how these payoffs were reached. This suggests that changes in information should have no impact on cooperation.

Proposition 2. *Intention-based reciprocity [8] predicts: (a) that second mover cooperation only depends on the first mover's choice. (b) With perfect information, conditional cooperation is only possible when the reversal probability is low. (c) Cooperation should be higher when information is imperfect.*

According to the intention-based model of reciprocity by Dufwenberg and Kirchsteiger [8] the utility of an individual i is as follows: $U_i = \pi_i + \lambda_i \cdot k_{ij} \cdot \phi_{iji}$ where $\lambda_i \geq 0$ and captures i's sensitivity towards reciprocity. The function k_{ij} captures person i's kindness towards person j. While the function ϕ_{iji} is a measure of i's belief about the kindness of j towards i. Both k_{ij} and ϕ_{iji} depend on individual first and second order beliefs (for details see Appendix A). This framework measures kindness at a particular node based on the difference between the resulting payoff and an equitable payoff. The equitable payoff is computed as the average of the maximum and minimum possible efficient payoffs. In other words, kindness is based on the current choice and what could have occurred if different choices were taken. This allows the intentions of others to matter.

With perfect information, cooperation in pure strategies by second movers is only possible when the reversal probability for the first mover's choice is low (see Proposition A2 in the Appendix A for details). In other words, the second mover will only cooperate when the first mover has a high degree of control over her actions. When control is high and with sufficient concern about reciprocity, the second mover will cooperate if the first mover cooperates and defect if the first mover defects. Importantly, according to this model, the

second mover will ignore the results of nature and condition their choice entirely on the first movers decision.

Under imperfect information, conditional cooperation is possible even when the reversal probability is high. As a result, second mover conditional cooperation may increase when the first mover's choice is uncertain (see Proposition A3 in the Appendix A for details). Caution must be taken with this result because it relies on the sequential reciprocity equilibrium holding in the imperfect information setting. Under this equilibrium concept, second movers know with probability one the choice of the first mover. This is a strong assumption that may not hold.

Both inequity aversion [6] and reciprocity [8] can give quite different predictions in the sequential prisoner's dilemma with nature. It is possible that people may care about both outcomes and intentions. In Appendix A, a mixed concerns model is developed that combines concerns for inequity aversion [6] and reciprocity [8] into a single framework. The model captures predictions of both models but suggests an additional equilibrium under perfect information where the second mover only cooperates if the both the first mover and nature cooperate.

Importantly, the mixed concerns model and other combined models of outcomes and intentions like Falk and Fischbacher [11] predict that cooperation should increase as the reversal probability decreases. Reciprocity in these models depends on the control a person has over her choices. When first movers have greater control, their decision to cooperate is viewed as a kinder action compared to when they have little control.

3.1. Individual Heterogeneity

3.1.1. Perspective Taking

The equilibrium concept assumed in the model of Dufwenberg and Kirchsteiger [8] is quite strong as it requires individuals to have correct higher order equilibrium beliefs. Empirical studies of individual beliefs suggest that this assumption may be too strict for some individuals. When subjects choose in both roles, Blanco et al. [29] found a "consensus effect" in a sequential prisoner's dilemma game where individual beliefs about the choices of others were biased towards one's own decision. In Dhaene and Bouckaert [18], both first and second order beliefs were similar in the sequential prisoner's dilemma but were biased in the ultimatum game. Individual differences in the ability to predict others behavior is one potential explanation for these results.

Predicting others behavior in strategic environments appears to depend on perspective taking [30,31]. Perspective taking is the ability to imagine or understand what another person is thinking or feeling [32]. When an individual engages in perspective taking they may "put themselves in another person's shoes." Evidence suggests that perspective taking develops as children age, is deficient in individuals who have autism, and is correlated with rule-following behavior [33–35].

In the sequential prisoner's dilemma with nature, first movers must use perspective taking to attempt to predict what potential second movers will do. Second movers may use perspective taking to try to understand the meaning behind the observed actions of the first mover. For second movers who have low perspective taking, assuming that they hold correct first and second order beliefs may be too strong. Low perspective taking may hold beliefs that are more likely to be biased. In the mixed concerns model, the equilibrium beliefs determine the kindness of the first mover's action. If a subject has low perspective taking then they may be more likely to misinterpret the meaning of others' actions. In the perfect information case, the first mover can signal their intended action via their choice of cooperation or defection. If control is low, cooperation by the first mover is a less costly signal. As a result, even subjects who do not have "good" intentions may choose to cooperate hoping that second movers will think that they have "good" intentions and subsequently reward them. Individuals who have high perspective taking should be more likely to recognize that selfish first movers may be more likely to cooperate when control is low. As a result, they should be less likely to reciprocally cooperate when control is low

compared to individuals who have low perspective taking. High perspective taking second movers may feel less guilty about defecting if the first mover cooperates because if control is low, then it is more likely they are defecting on a selfish cooperator. Similarly, when first mover control increases higher perspective taking should result in higher cooperation as they may be more likely to recognize the kindness of the first mover. When individuals have low perspective taking, observing higher cooperation by first movers may lead these second movers to think that people are being kind and as a result make these second movers more likely to cooperate when control is low.

Proposition 3. *Second movers with higher perspective taking should be more likely to conditionally cooperate as the first mover's control over their choices increases compared to those with lower perspective taking.*

3.1.2. Empathy

In the outcome-based and intention-based models of fairness it is assumed that each individual person can differ on how much they care about the different fairness concerns. While the models allow for individual heterogeneity it remains unclear why individuals differ in their concerns for fairness. One potential motivation for fair behavior may stem from individual capacity to empathize with others. In the *Theory of Moral Sentiments*, Adam Smith highlighted "compassion" or "fellow feeling" as the main factor in moral behaviors [36]. This factor is now known as empathy and is essential in order for humans to understand others [32].

Empathic concern is the feeling of compassion or concern an individual has for the welfare of another person and the desire to help [32,37]. In the empathy-altruism hypothesis, empathic concern is proposed as the motivation for altruistic behavior. In order for empathic concern to be activated, an individual must perceive that another person is in need or value the welfare of that person [38]. Once activated, empathic concern creates a desire to help that person. Empirical support suggests that empathic concern is an important component in altruistic behavior. Using survey evidence, empathic concern has been correlated with preferences for charitable giving [39,40], helping intentions [41], and distributive justice [40]. Additionally, empathic concern has been shown to be important in explaining behavior in dictator games [42,43], and public good games [44,45].

A few studies have suggested that empathic concern may be important in understanding cooperation in prisoner's dilemma games. In Batson and Moran [46], female subjects played a one-shot prisoner's dilemma game with one-way communication. In the communication treatment, female subjects thought they were receiving written communication from the other player but instead the experimenters sent a note describing a negative personal experience. To induce different levels of empathy for the other player, subjects were asked to read the note objectively (low empathy) or try to imagine the situation from the other person's point of view (high empathy). Although the sample size was small, cooperation was higher when subjects read the note and even higher when adopting the viewpoint of the other person. Rumble et al. [47] induced empathy in a similar way as Batson and Moran [46] in a repeated prisoner's dilemma. Subjects believed they were playing with another player, but they were actually playing with a pre-programmed computer. Treatments varied whether the computer played tit for tat (no noise), tit for tat with noise (noise), and a noncooperative strategy. The results showed that the high empathy condition sustained high cooperation in the noisy condition, but not in the noncooperative strategy. Batson and Ahmad [48] repeated the experiment in Batson and Moran [46] except that female subjects were told that the other player had defected. The results showed that only 5% of subjects chose to cooperate in the no empathy condition, but 45% cooperated in the high empathy condition. These studies suggest that higher empathic concern may increase cooperation, but they cannot explain whether empathic concern drives increased cooperation through positive reciprocity, distributional concerns, or a combination of both.

In order for empathic concern to be activated, the target must be perceived as in need of help. In the sequential prisoner's dilemma with nature, cooperation by the first mover requires trust that the second mover will cooperate as well. By cooperating, the first mover makes themselves more vulnerable to exploitation. People who have higher empathic concern could be more likely to cooperate if the first mover cooperates. If the first mover defects, then it is more likely that the second mover will receive a lower payoff. First mover defection could be viewed negatively by second movers, and subsequently not activate empathic concern. If empathic concern is not activated, then there should be no significant difference in cooperation rates based on empathic concern. This suggests that higher empathic concern should correlate with positive reciprocity, but not with negative reciprocity.

Proposition 4. *Second-movers with higher empathic concern should cooperate more if the first mover cooperated compared to those with lower empathic concern.*

4. Experiment

4.1. Experimental Design

A total of 246 students at a large public university participated in experimental sessions conducted in a computer laboratory. Students were recruited from a large subject pool. Recruitment to the subject pool took place through both classroom advertisements as well as through university emails. Prior to each experimental session, a random draw of students from the subject pool were sent an email about the upcoming session. Students then registered through the subject pool website. When registered students arrived at the experiment, they were randomly assigned to a computer terminal. No subject participated in more than one session.

The experiment was programmed and conducted with the software z-Tree [49]. At the start of the experiment, subjects were randomly assigned to the role of first or second mover. This role stayed the same throughout the experiment and each round subjects were randomly matched with a different subject. Due to not having exactly 40 subjects in each session there was some contamination in matching. After reading instructions, subjects completed a quiz to test comprehension of the instructions. A copy of the full instructions given to subjects can be found in the Supplementary Materials (See Figures S1–S12). Once finished, subjects played 20 rounds of a sequential prisoner's dilemma with nature. The role of nature was played by the computer. Each round the first mover could choose to cooperate or defect. After that choice, there was a random chance the computer reversed that choice. After both the first mover's choice and the results of the computer, the second mover then chose whether to cooperate or defect. To avoid potential framing effects subjects could choose A "cooperate" or B "defect."

Each subject participated in one of four possible treatments. Two sessions were conducted for each treatment for a total of eight sessions. The experiment used a within and between subjects design.

The reversal probability varied within subjects. Each treatment had a High Control condition and a Low Control condition. In the High Control condition, the chance the computer reversed the first movers choice was 10%.[4] For the Low Control condition, the reversal probability was 40%. Sessions were conducted where subjects received the High Control condition in the first ten rounds and the Low Control condition in the last ten rounds. Additionally, to control for order effects, sessions were conducted where subjects received the Low Control condition in the first ten rounds and the High Control condition in the last ten rounds. Prior to participating in the first ten rounds, each subject was told the reversal probability that would occur for those ten rounds. In order to allow between subject analysis, subjects were not told what the reversal probability would be in the last ten rounds until the start of the 11th round. The advantage of varying the control within subjects is that I can account for individual responses to the change in the reversal probability. Since it is assumed that utility functions vary by each individual i, using a

purely between subject design creates the concern in smaller samples that differences in subject responses could be due to random differences in the utility functions between subjects and not necessarily due to the treatment variables.

The information varied between subjects. Each subject participated in either the Known or Uncertain treatment. In the Known treatment, subjects were told what the first mover chose and what the computer chose. In the Uncertain treatment, subjects were only told what the computer chose.

Subjects received a $7 show up payment and were paid based on three randomly selected rounds. The experiment lasted an average of 40 min each session. Subjects could earn anywhere from $10 to $19. The average amount earned by subjects was approximately $14. Table 1 gives information about the demographic characteristics of subjects by treatment. The average age, number of economic classes, and number of statistics classes are quite similar across treatments. Overall, 61% of subjects were female.

Table 1. Treatment Information.

	High Control First		Low Control First		Total
	Known	Uncertain	Known	Uncertain	
Average:					
Age	20.34	20.53	20.56	20.24	20.42
Number of Economics Classes	1.18	1.33	1.26	1.53	1.32
Number of Statistics Classes	1.05	1.08	1.16	0.93	1.06
Take Home Earnings	13.66	13.83	13.76	13.84	13.77
Female (Fraction)	0.63	0.64	0.52	0.67	0.61
Number of Subjects	62	64	62	58	246

4.2. Questionnaire

After the experiment finished subjects completed a questionnaire containing demographic questions as well as psychometric tests designed to elicit levels of empathic concern and perspective taking. To measure empathic concern and perspective taking, I used a subset of the interpersonal reactivity index (IRI) [50]. The measure of empathic concern consisted of seven statements, for each statement subjects rated on a 5-point scale how well each statement described them. Examples of the statements are "I often have tender, concerned feelings for people less fortunate than me" and "When I see someone being treated unfairly, I sometimes don't feel very much pity." Similarly, the measure of perspective taking consisted of seven statements that subjects rated on a 5-point scale how well each statement described them. Examples of the statements are "I sometimes find it difficult to see things from the "other person's" point of view" and "When I'm upset at someone, I usually try to "put myself in that person's shoes". Both sets of statements for empathic concern and perspective taking had strong internal consistency with Cronbach's alphas of 0.73 and 0.70, respectively. Using subject responses, two variables representing empathic concern and perspective taking were derived using factor analysis. The eigenvalue for the empathic concern factor was 2.005 while the eigenvalue for the perspective taking factor was 2.009.

4.3. Hypotheses

Hypothesis 1a. *If subjects only care about inequity aversion, then there should be no differences in second mover cooperation as control or information changes.*

Hypothesis 1b. *If subjects care about reciprocity or mixed concerns, then cooperation rates should be larger in the Uncertain treatment relative to the Known treatment.*

Hypothesis 1a directly from Proposition A1 (see Appendix A for details). Changes in control and information do not change the outcomes at the end nodes of the game. As a

result outcome-based models like inequity aversion predict no change in cooperation rates across the treatments. Hypothesis 1b comes from the Propositions A2–A5 (see Appendix A for details). If individuals care about reciprocity, the distribution of these concerns are similar across the treatments, and first mover cooperation is similar as information changes, then overall second mover cooperation rates should be higher in the Uncertain treatment compared to the Known treatment.

Hypothesis 2a. *Concern for reciprocity predicts that second mover cooperation will only occur if the first mover cooperates.*

Hypothesis 2b. *Inequity aversion predicts that second mover cooperation should only occur if the computer cooperates.*

Hypothesis 2c. *If subjects have mixed concerns, then in addition to the reciprocity and inequity aversion predictions, the model also predicts an additional equilibrium in the High Control condition where subjects only cooperate if the first mover and nature cooperates.*

Hypothesis 2a follows from Proposition A2 while Hypothesis 2b follows from Proposition A1. Hypothesis 2c comes from Proposition A4 (see Appendix A for details), and only occurs in pure strategies in the High Control condition.

Hypothesis 3. *If the first mover cooperates, then both reciprocity and the mixed concerns model predict that second mover conditional cooperation will be larger in the High Control condition compared to the Low Control condition.*

This hypothesis results from Propositions A2 and A4 (see Appendix A for details). In the reciprocity model, first mover cooperation should be viewed as kinder by second movers in the High Control condition compared to the Low Control condition. If first movers cooperate, then the reciprocity and mixed concerns model predict that second mover cooperation should be higher in the High Control condition.

Hypothesis 4a. *In the Known treatment, second movers who have higher empathic concern should be more likely to cooperate if the first mover cooperated, but differences in empathic concern should have no effect on cooperation given the first mover defected.*

Hypothesis 4b. *In the Known treatment, second movers with higher perspective taking should be less likely to conditionally cooperate in the Low Control condition compared to the High Control condition.*

4.4. Potential Econometric Issues

Due to subjects making repeated choices in the experiment, the use of standard ordinary least squares regressions is problematic as it is unlikely that the independence assumption will be met. In addition, the decision to cooperate or defect is a binary variable suggesting that non-linear regression is more appropriate. To deal with these issues, this paper uses random effects probit regressions. The random effects probit model controls for random individual heterogeneity among subjects assuming there is no correlation between the individual error term and the independent variables. Although robust to other standard error assumptions, all regressions report clustered robust standard errors at the subject level. Where applicable, the Supplementary Materials contains additional robustness checks including lagged choice variables (Table S2), and fixed effects logit (Table S3). The fixed effect logit controls for subject specific effects that do not change over time.

Another issue is that both empathy and perspective taking were measured after subjects participated in the experiment. Experimental conditions may have influenced how subjects answered the empathy and perspective taking. If this is true, then any correlation

between the treatments and empathy and perspective taking could be due to the post elicitation of the measures. Table S1 and the subsequent analysis using nonparametric tests in the Supplementary Materials shows that scores on empathic concern and perspective taking were not significantly different from each other by treatment and condition. This suggests that the treatments themselves do not seem to have led subjects to answer differently to the empathy and perspective taking measures.

One potential concern is that due to possibly high correlation between empathy and perspective taking that including both these terms in the same regression may create multicollinearity. Additional regressions including only empathy or only perspective taking, show that the coefficient estimates are largely similar (See Tables S4–S7 in the Supplementary Materials). While there is high correlation between the two variables, the estimates for their effects on cooperation appear to be unaffected by their simultaneous inclusion in the regression.

5. Main Results

Result 1. *Second mover cooperation was higher in the Low Control and Uncertain condition.*

Figure 3 shows the average second mover cooperation by the computer choice. Second movers cooperated less often when the first movers choice was known compared to uncertain. This suggests that knowledge of the first movers choice mattered contrary to what is predicted by outcome-based fairness models. Table 2 gives results from random-effects probit regressions using the data from the first 10 rounds. The restriction allows a between subjects analysis, and the results suggest that second movers cooperated more often in the Uncertain treatment and Low Control condition. These effects seem to be additive since the interaction was not significant.[5] Interestingly, second movers were more likely to cooperate in the Uncertain treatment. Similar to Cox and Deck [28], it appears that subjects gave first movers the benefit of the doubt.

Result 2. *Inequity aversion, reciprocity, and mixed concerns are unable to fully explain the experimental results.*

Purely inequity averse second movers should cooperate only if the computer cooperates. Looking at the Known Treatment, Figure 4 shows the average second mover cooperation rates for the different paths of play. Clearly, cooperation was higher when the computer cooperated which is consistent with the predictions of inequity aversion. The results in Table 3 confirms that subjects were clearly drawn to the Pareto superior outcome. More specifically, as predicted by inequity aversion, cooperation appears to be significant only when the computer cooperated. However, in the Known treatment, second movers were more likely to choose the Pareto superior outcome if the first mover cooperated. Although, this result is only significant at the 10% level it is not predicted by inequity aversion. Purely reciprocal second movers in the Known treatment should cooperate only if the first mover cooperated and ignore the computers choice. The results in Table 3 show that when the first mover cooperated and the computer defected second mover cooperation was not significantly different from zero. Additionally, in contrast to predicted by reciprocity, second mover cooperation if the first mover cooperate was significantly different if the computer defected compared to if the computer cooperated.

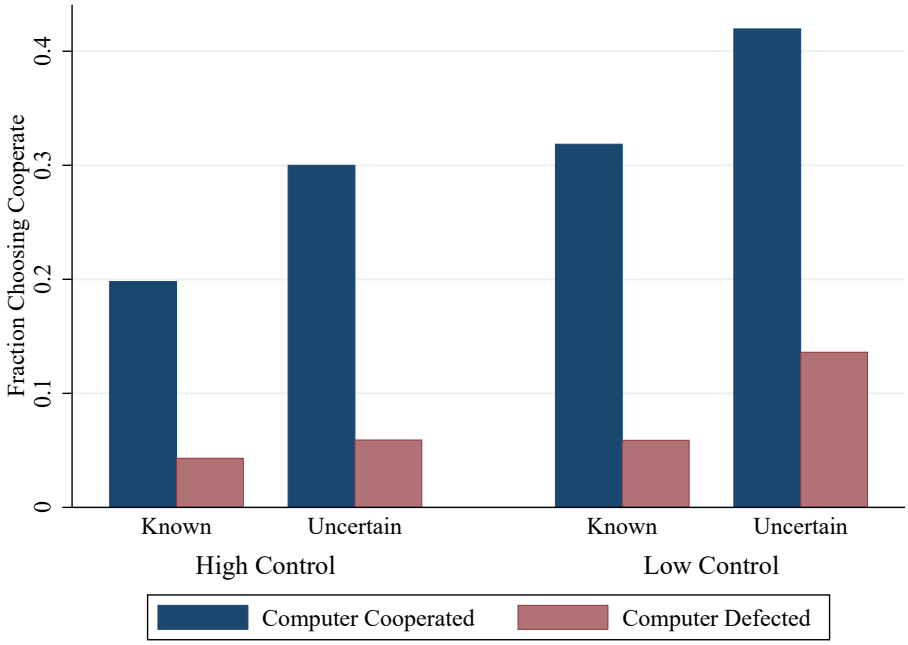

Figure 3. Average Second Mover Cooperation by Treatment (First 10 Rounds).

Table 2. First Mover and Second Mover Cooperation (First 10 Rounds).

	First Mover		Second Mover	
	Cooperation	**Predicted Probability**	**Cooperation**	**Predicted Probability**
Uncertain	−0.01	−0.00	0.47 *	0.06 *
	(0.13)	(0.04)	(0.21)	(0.03)
Low Control	0.42 **	0.15 **	0.46 *	0.06 *
	(0.13)	(0.04)	(0.21)	(0.03)
First Mover and			1.30 ***	0.26 ***
Computer cooperated			(0.15)	(0.04)
First Mover cooperated			0.10	0.01
and Computer defected			(0.24)	(0.03)
First Mover defected			1.20 ***	0.27 ***
and Computer cooperated			(0.17)	(0.05)
Female	0.10	0.04	−0.21	−0.03
	(0.13)	(0.05)	(0.21)	(0.03)
Intercept	−0.11		−1.52 ***	
	(0.18)		(0.27)	
N	1230	1230	1230	1230
ρ	0.23 ***	0.23 ***	0.45 ***	0.45 ***
Model χ^2	70.76	70.76	135.90	135.90

Predicted probabilities represent a discrete change from 0 to 1. Cluster robust standard errors at the subject level in parentheses. Results are from random-effects probit models that includes round fixed effects. Dependent variable is equal to 1 if player cooperated and equal to 0 otherwise. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Whenever the first mover's choice is reversed there are potentially competing norms for fair minded subjects. One hypothesis is that when faced with conflicting norms people will be more likely to select the norm that coincides with their own self-interest [51,52]. If first movers cooperated and the computer defected, cooperation by second movers was insignificant. While reciprocity suggests subjects should cooperate, inequity aversion predicts that subjects will defect. While this fits the hypothesis that people will select the norm that makes them personally better off, this does not seem to be the case when the first mover defected and computer cooperated. Here cooperation by second movers was significant. Reciprocity predicts that subjects should defect while inequity aversion

suggests people will cooperate. While there was less cooperation if the first mover defected, cooperation was much higher than we would suspect if people choose between competing norms by selecting the norm that maximizes their own payoff.

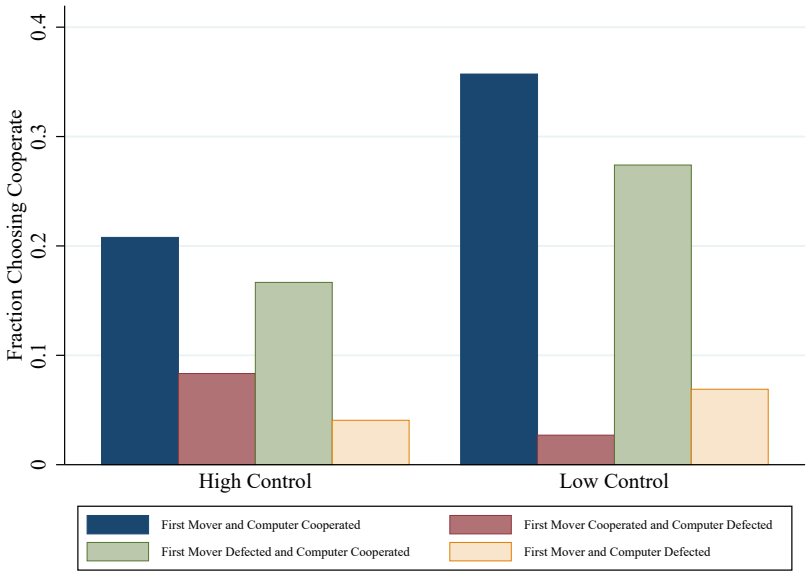

Figure 4. Average Second Mover Cooperation in Known Treatment (First 10 Rounds).

In the uncertain case, we fail to reject the predictions of inequity aversion but the reciprocity predictions are rejected. This test of the reciprocity model is more of a test of the SRE concept, because the concept requires that the second mover knows for certain that the first mover cooperated even though they only observed the computer defecting. This suggests that when players have imperfect information the SRE concept could be too strong of an assumption.

From Table 3 it appears that order mattered in the experiment. The variable Low Control First is equal to one if subjects received the Low Control condition in the first ten rounds. In both the Known and Uncertain treatments, subjects who started the experiment with Low Control cooperated at higher rates compared to subjects who received the Low Control second. This suggests that there was some path dependence in overall cooperation rates depending on which condition subjects received first. Despite this, the direction of change is consistent with the result that subjects cooperated more often when control was low compared to high.

Result 3. *In the Known treatment, if the first mover cooperated and the computer cooperated, second mover cooperation was higher in the Low Control condition compared to the High Control condition.*

Clearly, second movers were influenced by the different treatments and by the path of play. The mixed concerns model allows players to care about both what the first mover chose and the results from the computer. In the Known treatment, the mixed-concerns model predicts that when control is high and the first mover cooperates, then cooperation should be higher compared to the low control treatment. This prediction is not supported by the results. In round 1 of the Known treatment, given that both the first mover and the computer cooperated, second mover cooperation in the high control treatment was 17.6% compared to 52.9% in the low control treatment. These cooperation rates are significantly different from each other (Wilcoxon rank-sum test, $z = -2.121$, $p = 0.034$). This result is puzzling because theoretical predictions from reciprocity suggest that people should interpret cooperation by the first mover when control is high as kinder than cooperation when control is low. However, it appears that subjects responded in the opposite way as the model predicts. Since all end node payoffs for both players were kept constant, it

appears that the difference is primarily through how individuals were influenced by the reversal probability.

Table 3. Second Mover Cooperation by Treatment.

	Known		Uncertain	
	High Control	**Low Control**	**High Control**	**Low Control**
First Mover and	1.61 ***	1.84 ***	1.42 ***	1.50 ***
Computer cooperated	(0.25)	(0.28)	(0.23)	(0.29)
First Mover cooperated	−0.03	−0.32	0.07	0.33
and Computer defected	(0.63)	(0.46)	(0.64)	(0.33)
First Mover defected	1.02 **	1.42 ***	1.75 ***	1.57 ***
and Computer Cooperated	(0.31)	(0.29)	(0.33)	(0.27)
Low Control First	0.29	1.44 *	−0.16	2.11 **
	(0.50)	(0.61)	(0.53)	(0.69)
Female	−0.15	0.18	−0.63 +	−0.21
	(0.34)	(0.40)	(0.36)	(0.51)
Intercept	−1.74 ***	−3.01 ***	−1.19 **	−2.65 ***
	(0.44)	(0.62)	(0.44)	(0.70)
N	620	620	610	610
ρ	0.51	0.62	0.50	0.68
Model χ^2	59.61	66.98	61.45	65.73
Hypothesis Tests				
Inequity Aversion (Prob $> \chi^2(1)$)	0.07 +	0.07 +	0.31	0.78
Reciprocity (Prob $> \chi^2(1)$)	0.01 **	0.00 ***	0.04 *	0.00 ***

Cluster robust standard errors at the subject level in parentheses. Hypothesis for Inequity Aversion is that cooperation given computer cooperated is the same regardless of first mover's choice. Hypothesis for Reciprocity is that cooperation given first mover cooperated is the same regardless of computer's choice. Results are from random-effects probit regressions with round fixed effects. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Result 4. *(a) In the Known and Low Control Treatment, second movers with lower perspective taking were more likely to cooperate. (b) In the Known Treatment, second movers with higher empathic concern cooperated more often if the first mover cooperated.*

Perspective taking could be important in how others interpret or try to understand the actions of others. If individuals care about others' intentions, then perspective taking may be an influential factor in determining the intentions of others. Table 4 shows the results looking at the role of perspective taking on cooperation. In columns (1) and (2), differences in perspective taking do not appear to have influenced overall cooperation across all treatments. When the regressions were restricted to the Known Treatment, perspective taking by itself is not significant. However, when perspective taking is interacted with the Low Control condition the interaction is significant. Higher perspective taking was associated with lower cooperation rates in the Low Control treatment.

Figure 5 classifies individuals with lower than median perspective taking as low perspective and higher than median perspective taking as high perspective. In the high control treatment average conditional cooperation by high perspective takers is higher than low perspective takers. However, in the low control condition this is reversed. The regression results from Table 5 show that the interaction term is significant both when the first mover cooperated, and when the first mover defected. Figure 6 plots the predicted probabilities from Table 5 by scores in perspective taking for subjects in the High and Low Control conditions. When the first mover cooperated there is a clear decline in cooperation as scores in perspective taking increase. This decline actually crosses the high control treatment suggesting that individuals who have a high degree of perspective taking may have been more likely to cooperate in the High Control condition. Figure 7 plots the mean difference in predicted probability between the High and Low Control conditions by perspective taking. This graph shows that individuals who score low on perspective taking were significantly more likely to cooperate in the Low Control condition compared to low perspective takers in the High Control condition. While Figure 5 suggested that high

perspective takers may have been more likely to cooperate in the High Control condition this difference is not significant.

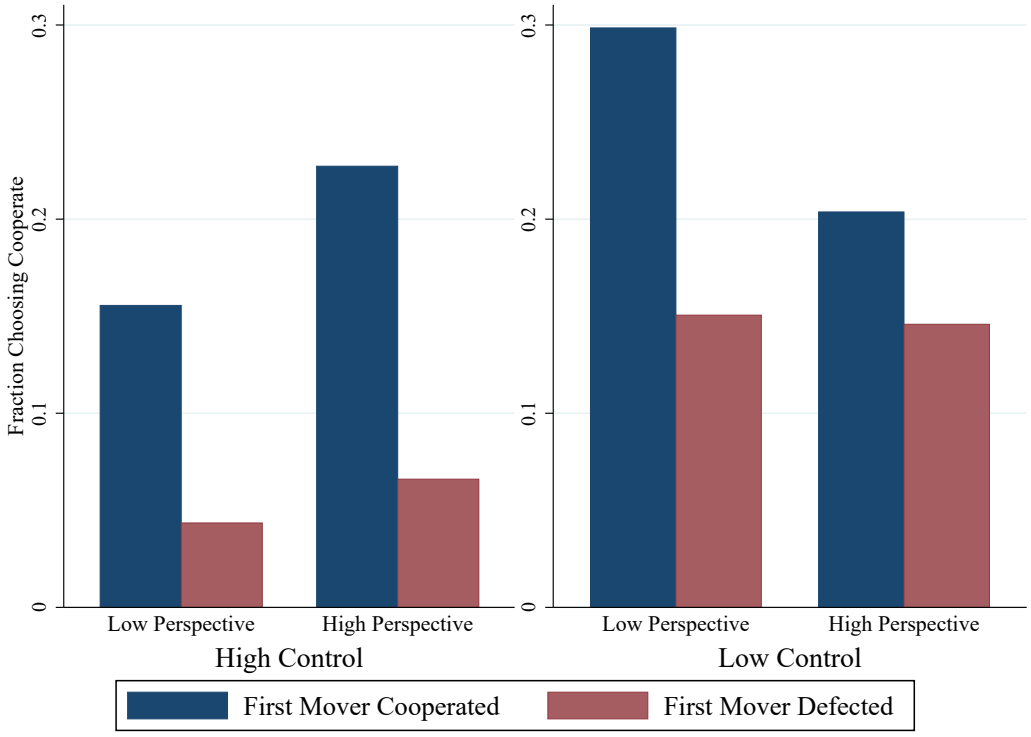

Figure 5. Average Second Mover Conditional Cooperation in Known Treatment by Perspective Taking (First 10 Rounds).

It is possible that subjects mistook first mover cooperation for kindness in the Low Control condition. One potential explanation for subjects with low perspective taking is that they just did not understand what they were doing. While this is possible, it seems unlikely because there was no effect from perspective taking in the High Control condition. One explanation for this could be that intentions were much easier to understand in this situation. When the first mover cooperated, most of the time the first movers choice matched the computer result. This may have made it easier for subjects to understand the meaning of the first movers choice despite differences in perspective taking abilities. Recent research has suggested that people have an automatic intuition to cooperate [53]. Low perspective takers may have found difficulty in inferring the meaning behind first mover cooperation in the Low Control condition. As a result, they may have been more likely to go with their gut instinct to cooperate.

Empathic concern could potentially lead to increased cooperation through it's influence on altruism. From Table 4, columns (1) and (2) support this view as subjects with higher empathic concern were more likely to cooperate overall. Columns (3) and (4) in Table 4 show that higher empathic concern is associated with increase cooperation in the known treatment. The coefficient for empathic concern is not significant when the regressions are restricted to the uncertain treatment. This suggests that higher empathic concern may only increase cooperation if subjects can view the actions of the first mover.

Table 5 examines the influence of empathic concern in the Known Treatment conditional on the first movers choice. If first movers cooperated, individuals with higher empathic concern were more likely to cooperate. If first movers defected, no significant difference in cooperation occurred based on differences in empathic concern. This supports the hypothesis that first mover cooperation activates empathic concern leading to cooperation. Defection by first movers does not activate empathic concern, and as a result

empathic concern does not influence behavior. It appears that empathic concern may be an important factor in individual desire to reward others for "good" intentions.

Table 4. Second Mover Cooperation with Empathy and Perspective Taking (First 10 Rounds).

	All Treatments		Known	
	(1) Second Mover Cooperation	(2) Second Mover Cooperation	(3) Second Mover Cooperation	(4) Second Mover Cooperation
Empathic Concern	0.30 *	0.31 *	0.41 $^{+}$	0.39 $^{+}$
	(0.13)	(0.13)	(0.23)	(0.22)
Perspective Taking	0.04	0.18	−0.21	0.14
	(0.13)	(0.15)	(0.23)	(0.28)
Low Control	0.55 *	0.56 *	0.45	0.38
	(0.21)	(0.22)	(0.34)	(0.33)
Low Control X Perspective Taking		−0.37		−0.87 **
		(0.24)		(0.34)
Uncertain	0.52 *	0.52 *		
	(0.21)	(0.21)		
First Mover and Computer cooperated	1.33 ***	1.33 ***	1.43 ***	1.42 ***
	(0.18)	(0.18)	(0.28)	(0.28)
First Mover cooperated and Computer defected	0.04	0.03	−0.38	−0.39
	(0.24)	(0.24)	(0.48)	(0.50)
First Mover defected and Computer Cooperated	1.23 ***	1.23 ***	1.14 ***	1.17 ***
	(0.19)	(0.19)	(0.28)	(0.28)
Female	−0.34	−0.32	−0.21	−0.07
	(0.21)	(0.22)	(0.34)	(0.31)
Intercept	−1.53 ***	−1.57 ***	−1.54 ***	−1.63 ***
	(0.29)	(0.29)	(0.47)	(0.46)
N	1200	1200	610	610
ρ	0.43	0.42	0.50	0.46
Model χ^2	137.51	138.86	64.17	65.65

Cluster robust standard errors at the subject level in parentheses. Regressions are restricted to the first 10 rounds. Results are from random-effects probit regressions with round fixed effects. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

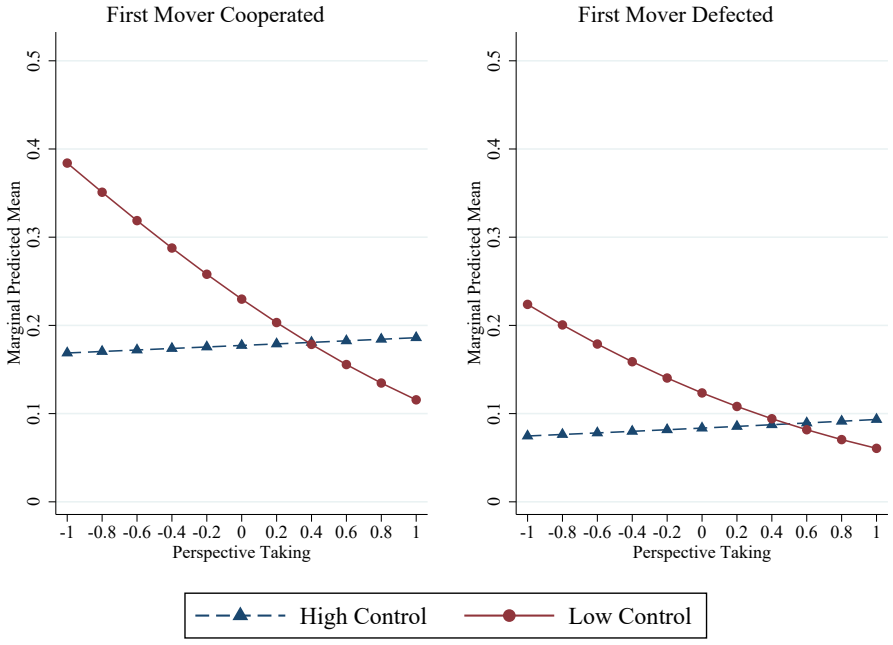

Figure 6. Probability by Perspective Taking in Known Treatment (First 10 Rounds).

Table 5. Conditional Cooperation in Known Treatment with Empathy and Perspective Taking.

	First Mover Cooperated	**First Mover Defected**
	Second Mover Cooperation	**Second Mover Cooperation**
Empathic Concern	0.64 *	0.25
	(0.31)	(0.21)
Perspective Taking	0.07	0.09
	(0.41)	(0.23)
Low Control	0.37	0.34
	(0.53)	(0.30)
Low Control X	−0.97 *	−0.68 *
Perspective Taking	(0.48)	(0.34)
Computer cooperated	2.18 *	1.14 ***
	(0.89)	(0.27)
Female	−0.16	−0.10
	(0.47)	(0.36)
Intercept	−2.77 *	−0.90 +
	(1.13)	(0.50)
N	209	306
ρ	0.61	0.34
Model χ^2	21.95	29.68

Cluster robust standard errors at the subject level are in parentheses. Regressions are restricted to the first 10 rounds. Results are from random-effects probit regressions with round fixed effects. $^{+}$ $p < 0.10$, * $p < 0.05$, *** $p < 0.001$.

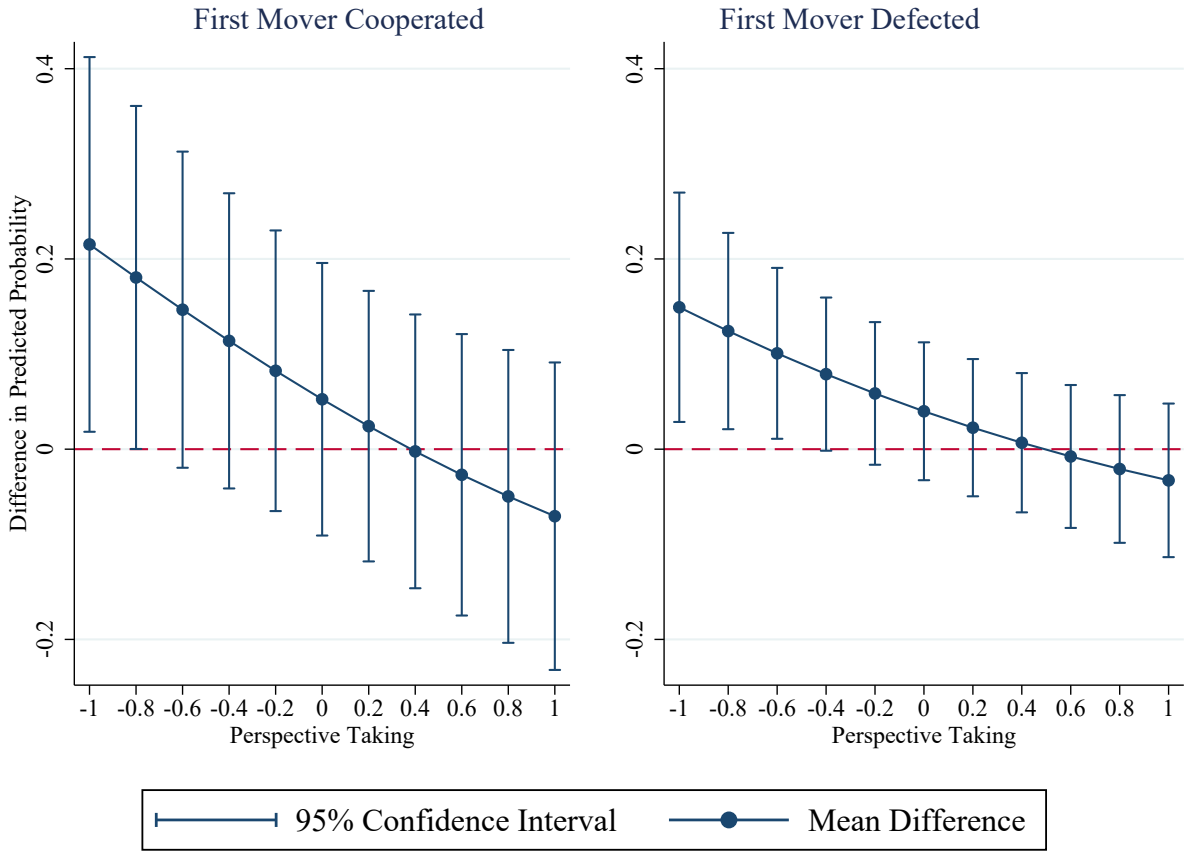

Figure 7. Mean Difference in Predicted Probability between Low Control and High Control conditions by Perspective Taking (First 10 Rounds and Known Treatment).

6. Additional Results

While the focus of the paper is on second mover behavior, additional insights in understanding cooperation in the sequential prisoner's dilemma can be learned from examining first mover behavior. Figure 8 shows the average first mover cooperation by treatment restricted to the first 10 rounds. First movers cooperated more often in the Low Control condition compared to the High Control Condition. Cooperation between the Known and Uncertain treatments appears to be similar. Table 2 shows that cooperation by the first mover was significantly higher in the Low Control condition. Higher cooperation by first movers when control is low could be due to the fact that cooperation is a less costly signal since there is a high chance that their choice will be reversed. In the uncertain treatment, first movers could avoid having subjects learn about their choices, but they were still aware of how their choice influenced the second movers. This is similar to the "plausible deniability" treatment in the dictator game experiment by Dana et al. [54]. Interestingly, first mover behavior was not influenced by whether their intended choice would be known or unknown to the second mover.

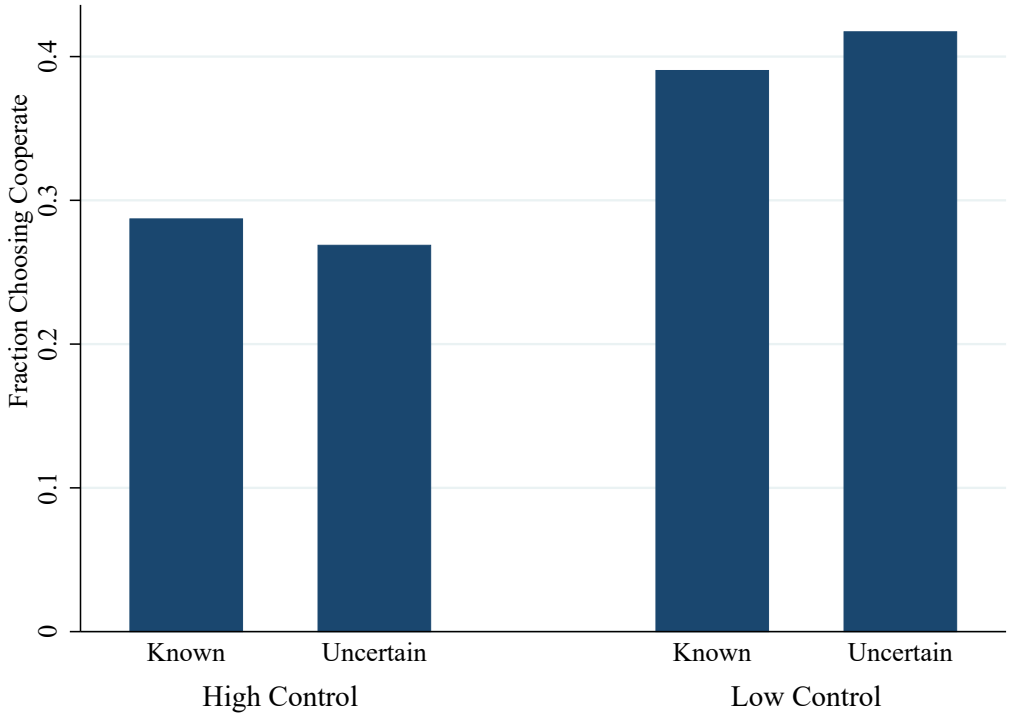

Figure 8. Average First Mover Cooperation by Treatment (First 10 Rounds).

7. Conclusions

This paper has shown that cooperation rates in the sequential prisoner's dilemma with nature are greater than predicted by pure self-interest. The failure of pure-self interest to explain behavior in a wide range of one-shot experimental games has led many researchers to suggest fairness concerns as a potential explanation for the empirical results [55]. Theoretical models of fairness can be classified into two types: outcome-based and intention-based. Outcome-based models of fairness assume that people care about fair distributions [5,6]. These models assume that intentions are not relevant for predicting behavior. The results from this experiment suggest that outcomes matter, but purely consequentialist models cannot fully explain subject behavior. Intention-based models capture the idea that people are reciprocal, preferring to be kind to people who are kind to them and punish people who are unkind to them [7,8]. In this experiment intentions mattered as well, but current models of intentions failed to explain the concerns for fair outcomes.

The mixed motives model combined both inequity aversion and reciprocity into a single framework. While able to account for concerns about both outcomes and intentions, the model was unable to explain the increased second mover cooperation in the Low Control treatment. The results highlight that how individuals perceive others' intentions is still an open question. It is not possible for people to know with certainty the true intentions of another person. Despite this, people potentially use the observed actions of others to infer intentions. These intentions could influence how people respond to others' behavior and appear to be important in understanding cooperation.

One important aspect for understanding the attribution of intentions may be perspective taking. Differences in the ability to take the viewpoint of others was shown to be important in explaining increased cooperation by second movers when control was low. While concerns for fairness are important, both empathic concern and perspective taking could be significant factors in explaining subject behavior in games. Second movers who scored higher on empathic concern where more likely to cooperate. When the first movers choice was known, higher empathic concern increased cooperation when the first mover cooperated but not when the second mover defected. This demonstrates that empathic concern is different from negative reciprocity, since subjects with higher empathic concern were not more likely to defect if the first mover defected. Instead, empathic concern is thought to motivate altruism. This altruistic motivation is activated when the first mover cooperates leading to increased cooperation. Future research should include measures of empathic concern and perspective taking to investigate potential relationship in other games and environments.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/g12030058/s1, Figure S1: Experimental Instructions, Figure S2: Experimental Instructions-continued, Figure S3: Experimental Instructions- continued, Figure S4: Experimental Instructions-Known Treatment Only, Figure S5: Experimental Instructions- Uncertain Treatment Only, Figure S6: Experimental Instructions- Payoff Table, Figure S7: Experimental Instructions- Comprehension Questions, Figure S8: Experimental Instructions- Comprehension Answers, Figure S9: Experimental Instructions- Known Treatment Only, Figure S10: Experimental Instructions- Uncertain Treatment Only, Figure S11: Experimental Instructions Part II- Known Treatment Only, Figure S12: Experimental Instructions Part II- Uncertain Treatment Only, Table S1: Empathy and Perspective Taking Summary Statistics, Table S2: Second Mover Cooperation by Treatment with lagged variables, Table S3: Second Mover Cooperation by Treatment- Logit Regressions, Table S4: Second Mover Cooperation with Empathy and Perspective Taking Robustness(First 10Rounds), Table S5: Second Mover Conditional Cooperation with Empathy and Perspective Taking Robustness, Table S6: Second Mover Conditional Cooperation in Known Treatment with Empathy- Robustness, Table S7: Second Mover Conditional Cooperation in Known Treatment with Perspective Taking-Robustness.

Funding: This research received no external funding.

Institutional Review Board Statement: This project was approved by the University of California, Irvine Institutional Review Board under protocol HS#2011-8378.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Theoretical Models of Social Preferences

Appendix A.1. Inequity Aversion

Proposition A1. *For player 2, in any pure strategy subgame perfect Nash equilibrium:*

1. *If* $\beta_2 > \frac{1}{3}$, *then player 2 will cooperate if nature cooperates and defect if nature defects.*
2. *If* $\beta_2 < \frac{1}{3}$, *then player 2 will always defect.*

Proof. To see that 1 holds, first let us look at when player 1 and nature cooperates. Player 2 will cooperate if $3 > 4 - 3 \cdot \beta_2$. This occurs when $\beta_2 > \frac{1}{3}$. Due to symmetry, this condition also ensures that if player 1 defects and nature cooperates, then player 2 will cooperate. To see that player 2 will defect if nature defects, let us look at the case when player 1 cooperates and nature defects. Player 2 will choose to defect if $2 > 1 - 3 \cdot \alpha_2$ which holds for all α_2 since $\alpha_2 \in [0, 1]$. Due to symmetry we can see that this will also hold if player 1 defects and nature defects. □

Appendix A.2. Reciprocity Model

In this section, I introduce the intention-based reciprocity model of Dufwenberg and Kirchsteiger [8]. The model uses psychological game theory based on Battigalli and Dufwenberg [56]. Psychological games, first developed by Geanakoplos et al. [57], differ from standard games in that an individual's beliefs directly affects her utility. Using psychological game theory, Rabin [7] modeled concerns for reciprocity in normal form games and Dufwenberg and Kirchsteiger [8] extended the idea to extensive form games. Reciprocity is captured by the idea that people like to be kind to people who are kind to them and be unkind to people who are unkind to them.

Formally, let $I \in \{0, 1, 2\}$ be the set of players. Denote nature as player 0. Let H be the set of histories that lead to subgames. Each player $i \in I \setminus \{0\}$ has a set of possible strategies A_i. The strategy set is $A = \prod_{i \in I \setminus \{0\}} A_i$. Each strategy $a_i \in A_i$ gives a probability distribution on the possible choices of player i at each history $h \in H$. Each player $i's$ updated strategy is defined as $a_i(h)$.[6] The probability distribution for the behavioral strategy of the chance player is defined as θ which is commonly known to both players. Given end node payoffs, the expected material payoff for each player $i \in I \setminus \{0\}$ is $\pi_i : A \times \{\theta\} \to \Re$.

Following Dufwenberg and Kirchsteiger [8] and Sebald [58], additional notation must be introduced as it is necessary to keep track of first and second order beliefs. Each player i has a set of beliefs, B_{ij}, about the strategy of player j. Let $b_{ij} \in B_{ij}$ be the belief player i has about the strategy of player j. Let C_{iji} define the set of beliefs player i has about the belief player j has about player $i's$ strategy. Define $c_{iji} \in C_{iji}$ as the belief player i has about the belief player j has about player $i's$ strategy. To capture the main features of reciprocity beliefs need to be updated as the game progresses. Let $b_{ij}(h)$ and $c_{iji}(h)$ represent the updated beliefs at history h. The utility for player i is defined as follows:

$$\begin{aligned} U_i(a_i(h), b_{ij}(h), c_{iji}(h), \theta) &= \pi_i(a_i(h), a_j(h), \theta) \\ &+ \lambda_i \cdot k_{ij}(a_i(h), b_{ij}(h), \theta) \cdot \phi_{iji}(b_{ij}(h), c_{iji}(h), \theta)] \end{aligned} \tag{A1}$$

where $\lambda_i \geq 0$.

In A.1, $i's$ utility depends on $i's$ own payoff plus concerns for reciprocity. The weight that i places on reciprocity concerns is captured by λ_i. The function $k_{ij}(a_i(h), b_{ij}(h), \theta)$ is a measure of the kindness of i towards j at history h, and $\phi_{iji}(b_{ij}(h), c_{iji}(h), \theta)$ is $i's$ belief about the kindness of j towards i at history h. The kindness of player i towards player j is represented as a function of player $i's$ strategy choice $a_i(h)$ and belief $b_{ij}(h)$. At a specific $a_i(h)$ and $b_{ij}(h)$ the kindness of player i is captured by the payoff that player j gets minus an equitable payoff. The kindness function is defined as:

$$k_{ij}(a_i(h), b_{ij}(h), \theta) = \pi_j(a_i(h), b_{ij}(h), \theta) - \pi_j^{ei}((b_{ij}(h), \theta) \tag{A2}$$

Dufwenberg and Kirchsteiger [8] calculate the equitable payoff for player j by finding the a_i that gives player j the highest possible payoff and finding the a_i that gives player j the lowest possible payoff. The equitable payoff is an average of the payoffs for player j evaluated at each a_i. This equitable payoff is:

$$\pi_j^{ei}(b_{ij}, \theta) = \frac{1}{2}[\max_{a_i \in A_i}\{\pi_j(a_i, b_{ij}, \theta)\} + \min_{a_i \in E_i}\{\pi_j(a_i, b_{ij}, \theta)\}] \tag{A3}$$

where E_i, defined by Dufwenberg and Kirchsteiger [8], is the set of efficient strategies for player i such that

$$E_i = \{a_i \in A_i | \text{ there exists no } a_i' \in A_i \text{ such that for all } h \in H, (a_j)_{j\neq i} \in \prod_{j\neq i} A_j, \text{ and } k \in I \text{ it holds that } \pi_k(a_i'(h), (a_j(h))_{j\neq i}) \geq \pi_k(a_i(h), (a_j(h))_{j\neq i}), \text{ with strict inequality for some } (h, (a_j)_{j\neq i}, k)\}. \quad \text{(A4)}$$

Player $i's$ belief about the kindness of player j towards player i has a similar structure and is defined as[7]:

$$\phi_{iji}(b_{ij}(h), c_{iji}(h), \theta) = \pi_i(b_{ij}(h), c_{iji}(h), \varepsilon) - \pi_j^{ej}(c_{iji}(h), \theta) \quad \text{(A5)}$$

The equilibrium concept used in this paper is the sequential reciprocity equilibrium [8,58]. Define for all $a = (a_i)_{i\in I} \in A$ and history $h \in H$, let $A_i(a, h) \subseteq A_i$ be the set of behavioral strategies for each player i that give the same choices as the strategy $a_i(h)$ for all histories other than h.

Definition A1. *The profile $a^\star = (a_i^\star)_{i\in I\setminus\{0\}}$ is a sequential reciprocity equilibrium (SRE) if for all $i \in I \setminus \{0\}$ and for each history $h \in H$ the following properties hold:*
(i) $a_i^\star(h) \in \underset{a_i\in A_i(h,a)}{argmax}\, U_i(a_i(h), b_{ij}(h), c_{iji}(h), \theta)$ *where* $i \neq j$
(ii) $b_{ij} = a_j^\star$ *for all* $j \neq i$
(iii) $c_{iji} = a_i^\star$ *for all* $j \neq i$

Property (i) means that at history h, player i chooses a strategy profile that maximizes $i's$ utility given $i's$ belief. In addition, it assures that player i follows the equilibrium strategy at all other histories. At the initial history, properties (ii) and (iii) imply that initial beliefs are correct. Property (i) adds that any sequence of choices that lead to a history have probability one. As a result, the SRE concept requires that in equilibrium beliefs be correct.

Proposition A2. *Under perfect information and if $\theta < \frac{1}{2}$, then in any SRE the potential behavior for player 2 can be described as follows:*

(a) *If $\theta < \frac{1}{4}$, and $\lambda_2 > \frac{1}{1-4\theta}$, then player 2 will cooperate if player 1 cooperates and defect if player 1 defects.*
(b) *$\lambda_2 < \frac{1}{2-4\theta}$, then player 2 will always defect.*

Proof. Player 2 can choose to cooperate or defect at each node h^3, h^4, h^5, and h^6 labeled in Figure 1. Let player 1's belief about what player 2 will choose at each node be defined as: $x_1 = P_1(2 \text{ choses } C|h^3)$, $x_2 = P_1(2 \text{ choses } C|h^4)$, $x_3 = P_1(2 \text{ choses } C|h^5)$, and $x_4 = P_1(2 \text{ choses } C|h^6)$. Player 2's belief about player 1's belief about what player 2 will choose at each node is defined as the expectation of player 1's beliefs about player 2. This gives: $y_1 = E_2[x_1|h^3]$, $y_2 = E_2[x_2|h^4]$, $y_3 = E_2[x_3|h^5]$, and $y_4 = E_2[x_4|h^6]$. Player 1 can choose to cooperate or defect at node h^0. Let player 2's belief that player 1 will cooperate be $z_1 = P_2(1 \text{ choses } C|h^0)$. Player 1's belief about player 2's belief that player 1 will cooperate is defined as $w_1 = E_1[z_1|h^0]$. The game can now be analyzed as a psychological game with reciprocity.

If player 1 cooperates, then player 2's belief about the the kindness of player 1 towards player 2 is $\phi_{212} = \frac{1}{2}((1-\theta)(4-y_1) + \theta(2-y_2) - \theta(4-y_3) - (1-\theta)(2-y_4))$. If player 1 defects, then $\phi_{212} = \frac{1}{2}(\theta(4-y_3) + (1-\theta)(2-y_4) - (1-\theta)(4-y_1) - (\theta)(2-y_2))$. In any SRE player 2 will always make the same decision at history h^5 and h^6. To see why, note that for player 2 to defect at h^6 it must be that $1 + \lambda_2[(1-\theta)(4-y_1) + \theta(2-y_2) - \theta(4-y_3) - (1-\theta)(2-y_4)] > 0$. In order for the second mover to defect at h^5 it must be that $1 + \lambda_2[(1-\theta)(4-y_1) + \theta(2-y_2) - \theta(4-y_3) - (1-\theta)(2-y_4)] > 0$. As a result, in any

SRE it must be the case that $x_3 = x_4 = y_3 = y_4$. Similarly, in any SRE player 2 will always make the same decision at history h^1 and h^2.

If (a) holds in equilibrium, then $x_1 = x_2 = y_1 = y_2 = 1$ and $x_3 = x_4 = y_3 = y_4 = 0$. If player 1 cooperates, then $\phi_{212} = \frac{1}{2}(1 - 4\theta)$. At h^3 player 2 will cooperate if $3 + (\frac{1}{2})\lambda_2(1 - \gamma_2)(1 - 4\theta) > 4 - (\frac{1}{2})\lambda_2(1 - \gamma_2)(1 - 4\theta)$. This holds if $\theta < \frac{1}{4}$ and $\lambda_2 > \frac{1}{1-4\theta}$. At h^4, player 2 will cooperate if $\lambda_2(1 - 4\theta) > 1$. This holds if $\theta < \frac{1}{4}$ and $\lambda_2 > \frac{1}{1-4\theta}$. Since $\gamma_2 \in [0, 1]$, then in order for player 2 to cooperate in pure strategies at h^4 it must be the case that $\theta > \frac{1}{4}$. For player 2 to defect at h^5, then the following must hold $\lambda_2(1 - 4\theta) > -1$. Since $\lambda_2(1 - 4\theta) > 1$, then player 2 will defect at h^5. As a result, if $\theta > \frac{1}{4}$, and $\lambda_2 > \frac{1}{1-4\theta}$, then player 2 will cooperate if player 1 cooperates and defect if player 1 defects.

For (b) it must be the case that $y_1 = y_2 = y_3 = y_4 = 0$. If player 1 cooperates, then $\phi_{212} = \frac{1}{2}(2 - 4\theta)$. At h^3 player 2 will defect if $\lambda_2(2 - 4\theta) + < 1$. This holds if $\lambda_2 < \frac{1}{2-4\theta}$. At h^4 player 2 will defect if $\lambda_2(2 - 4\theta) < 1$ which holds if $\lambda_2 < \frac{1}{2-4\theta}$. At h^5 player 2 will defect if $\lambda_2(2 - 4\theta) > -1$. As a result, if $\lambda < \frac{1}{2-4\theta}$, then player 2 will always defect. □

Proposition A3. *Under imperfect information and if $\theta < \frac{1}{2}$, then in any SRE the potential behavior for player 2 can be described as follows:*

1. *If player 1 cooperates*
   - (a) $\lambda_2 > \frac{1}{2-4\theta}$, *then player 2 will always cooperate.*
   - (b) $\lambda_2 < \frac{1}{2-4\theta}$, *then player 2 will always defect.*
2. *If player 1 defects, then player 2 will always defect.*

Proof. Let player 1's belief about what player 2 will choose at each information set be defined as: $q_1 = P_1(2\ choses\ C|h^3 \cup h^5)$, and $q_2 = P_1(2\ choses\ C|h^4 \cup h^6)$. Player 2's belief about player 1's belief about what player 2 will choose at each information set is defined as the expectation of player 1's beliefs about player 2. This gives: $v_1 = E_2[q_1|h^3 \cup h^5]$, and $v_2 = E_2[q_2|h^4 \cup h^6]$. Player 1 can choose to cooperate or defect at node h^0. Let player 2's belief that player 1 will cooperate be $z_1 = P_2(1\ choses\ C|h^0)$. Player 1's belief about player 2's belief that player 1 will cooperate is defined as $w_1 = E_1[z_1|h^0]$.

Player 2 only observes the results of nature. Player 2's evaluation of the kindness of player 1 depends on the belief about what node she is currently at. If player 2 observes cooperation, then the probability that player 2 believes she is at node h^3 is $P(h^3|2\ observes\ C) = \frac{z_1 \cdot \varepsilon_1}{z_1 \cdot \varepsilon_1 + (1-z_1) \cdot \varepsilon_2}$ via Bayes rule. Similarly, $P(h^5|2\ observes\ C) = \frac{(1-z_1) \cdot \varepsilon_2}{z_1 \cdot \varepsilon_1 + (1-z_1) \cdot \varepsilon_2}$. If player 2 observes defection, then $P(h^4|2\ observes\ D) = \frac{z_1 \cdot (1-\varepsilon_1)}{z_1 \cdot (1-\varepsilon_1) + (1-z_1) \cdot (1-\varepsilon_2)}$ and $P(h^6|2\ observes\ D) = \frac{(1-z_1) \cdot (1-\varepsilon_2)}{z_1 \cdot (1-\varepsilon_1) + (1-z_1) \cdot (1-\varepsilon_2)}$. Since the SRE concept requires that initial beliefs be correct, it follows that player 2 knows in equilibrium what player 1 chooses.

Player 2's belief about the kindness of player 1 is $\phi_{212} = (z_1 - \frac{1}{2})(1 - 2\theta)(2 - v_1 + v_2)$. No matter the history, the kindness of player 2 towards player 1 will always be $k_{21} = 1$ if player 2 cooperates and $k_{21} = -1$ if player 2 defects.

Suppose that in equilibrium player 1 cooperates, then player 2 knows this. Since $z_1 = w_1 = 1$, player 2 knows that if nature cooperates then she is at node h^3 and if nature defects then she is at node h^4. If nature cooperates, then player 2 will cooperate if $\lambda_2(1 - 2\theta)(2 - v_1 + v_2) > 1$. If nature defects, then player 2 will cooperate if $\lambda_2(1 - 2\theta)$ $(2 - v_1 + v_2) > 1$. As a result, player 2 will make the same decision at nodes h^3 and h^4. Similarly, if in equilibrium player 1 defects, $z_1 = w_1 = 0$. In this case, player 2 will make the same decision at nodes h^4 and h^5.

For 1(a), $v_1 = v_2 = 1$. This is possible if $\lambda_2(1 - 2\theta)(2) > 1$. If $\lambda_2 > \frac{1}{2-4\theta}$, then player 1 will cooperate no matter the results of nature.

For 1(c), $v_1 = v_2 = 0$. This is possible if both $\lambda_2(1 - 2\theta)(2) < 1$. If $\lambda_2 < \frac{1}{2-4\theta}$, then player 2 will always defect.

For 2, $v_1 = v_2 = 0$. This is possible if $-\lambda_2(1-2\theta)(2) < 1$, which holds for all $\lambda_2 \geq 0$. As a result, if player 1 defects, then player 2 will defect. □

Appendix A.3. Mixed-Concerns Model

In this section, I introduce the mixed-concerns model that combines the models of inequity aversion [6] and reciprocity [8] into a single framework. Let the utility of an individual i be defined as:

$$\begin{aligned} U_i(a_i(h), b_{ij}(h), c_{iji}(h), \theta) &= \pi_i(a_i(h), a_j(h), \theta) \\ +\rho_i \cdot [(1-\gamma_i) \cdot k_{ij}(a_i(h), b_{ij}(h), \theta) \cdot \phi_{iji}(b_{ij}(h), c_{iji}(h), \theta) &+ \gamma_i \cdot D_{ij}(a_i(h), b_{ij}(h), \theta)] \end{aligned} \tag{A6}$$

where $\rho_i \geq 0$ and $\lambda_i \in [0,1]$. In Equation (A6), $i's$ utility depends on $i's$ own payoff plus concerns for reciprocity and inequity aversion. The weight that i places on these social preferences is captured by ρ_i. An additional parameter, γ_i, is the relative weight placed on concerns for reciprocity and distribution. Higher values of γ_i mean that person i places a lower weight on reciprocity and greater weight on distributional concerns.

The function $D_{ij}(a_i(h), b_{ij}(h), \theta)$ captures the distributional concerns of an individual. $D_{ij}(a_i(h), b_{ij}(h), \theta)$ is assumed to be a modified version of inequity aversion defined as:

$$D_{ij}(a_i(h), b_{ij}(h), \varepsilon) = -\max\{\pi_j - \pi_i,\ \pi_i - \pi_j\}$$

where $\pi_j = \pi_j(a_i(h), b_{ij}(h), \theta)$ and $\pi_i = \pi_i(a_i(h), b_{ij}(h), \theta)$.[8] The functional form for $D_{ij}(a_i(h), b_{ij}(h),)$ does not capture the idea from the inequity aversion model that people might dislike getting less than another person more than they feel bad about getting more. This could easily be incorporated into the model, but has been left out for simplicity.[9]

The function $k_{ij}(a_i(h), b_{ij}(h), \theta)$ is a measure of the kindness of i towards j at history h, and $\phi_{iji}(b_{ij}(h), c_{iji}(h), \theta)$ is $i's$ belief about the kindness of j towards i at history h. Both $k_{ij}(a_i(h), b_{ij}(h), \theta)$ and $\phi_{iji}(b_{ij}(h), c_{iji}(h), \theta)$ have the same functional form described in the previous section. Since the focus is on sequential games, the analysis uses the sequential reciprocity equilibrium as it allows beliefs to be updated.

One advantage of the mixed-concerns model compared to other models that combine concerns for intentions and outcomes [10,11], is that chance players are incorporated into the model. Chance players are often used in theoretical models to capture many different environments and random devices are often used in experiments. The mixed-concerns model can make equilibrium predictions in these situations. In addition, the model allows us to investigate how changes in the distribution of the choices by chance players influences equilibrium predictions.

Appendix A.3.1. Perfect Information

The main focus on this analysis will be on what player 2 chooses to do in the game with perfect information. In any SRE, the potential behavior for player 2 is described in Proposition 1.

Proposition A4. *If $\theta < \frac{1}{2}$, then in any SRE the potential behavior for player 2 can be described as follows:*

(a) *If $\theta < \frac{1}{4}$, $0 \leq \gamma_2 \leq \frac{1-4\theta}{4-4\theta}$, and $\rho_2 > \frac{1}{1-4\theta-\gamma_2(4-4\theta)}$, then player 2 will cooperate if player 1 cooperates and defect if player 1 defects.*

(b) *If $\theta < \frac{1}{3}$, $0 < \gamma_2 < \frac{1-3\theta}{4-3\theta}$, and $\frac{1}{1-3\theta+\gamma_2(2+2\theta)} < \rho_2 < \frac{1}{1-3\theta-\gamma_2(2-4\theta)}$, then player 2 will cooperate if player 1 and nature cooperates, and defect otherwise.*

(c) *If* $\frac{1-2\theta}{4-2\theta} < \gamma_2$ *and* $\rho_2 > \frac{1}{\gamma_2(2+2\theta)-1+2\theta}$, *then player 2 will cooperate if nature cooperates and defect if nature defects.*

(d) *If* $\gamma_2 > \frac{2-4\theta}{5-4\theta}$ *and* $\rho_2 < \frac{1}{2-4\theta+\gamma_2(4\theta-1)}$, *or* $\gamma_2 < \frac{2-4\theta}{5-4\theta}$ *and* $\rho_2 < \frac{1}{2-4\theta+\gamma_2(5-4\theta)}$, *then player 2 will always defect.*

Proof. Player 2 can choose to cooperate or defect at each node h^3, h^4, h^5, and h^6 labeled in Figure 1. Let player 1's belief about what player 2 will choose at each node be defined as: $x_1 = P_1(2 \textit{ choses } C|h^3)$, $x_2 = P_1(2 \textit{ choses } C|h^4)$, $x_3 = P_1(2 \textit{ choses } C|h^5)$, and $x_4 = P_1(2 \textit{ choses } C|h^6)$. Player 2's belief about player 1's belief about what player 2 will choose at each node is defined as the expectation of player 1's beliefs about player 2. This gives: $y_1 = E_2[x_1|h^3]$, $y_2 = E_2[x_2|h^4]$, $y_3 = E_2[x_3|h^5]$, and $y_4 = E_2[x_4|h^6]$. Player 1 can choose to cooperate or defect at node h^0. Let player 2's belief that player 1 will cooperate be $z_1 = P_2(1 \textit{ choses } C|h^0)$. Player 1's belief about player 2's belief that player 1 will cooperate is defined as $w_1 = E_1[z_1|h^0]$. The game can now be analyzed as a psychological game with mixed concerns.

If player 1 cooperates, then player 2's belief about the the kindness of player 1 towards player 2 is $\phi_{212} = \frac{1}{2}((1-\theta)(4-y_1) + \theta(2-y_2) - \theta(4-y_3) - (1-\theta)(2-y_4))$. If player 1 defects, then $\phi_{212} = \frac{1}{2}(\theta(4-y_3) + (1-\theta)(2-y_4) - (1-\theta)(4-y_1) - (\theta)(2-y_2))$. In any SRE player 2 will always defect at history h^6. To see why, note that for player 2 to defect at h^6 it must be that $1 + \rho_2(1-\gamma_2)[(1-\theta)(4-y_1) + \theta(2-y_2) - \theta(4-y_3) - (1-\theta)(2-y_4)] + 3\rho_2\gamma_2 > 0$. This holds if $\rho_2 \geq 0$ and $\gamma_2 \geq 0$. As a result, in any SRE it must be the case that $y_4 = x_4 = 0$. In addition, player 2 will not cooperate at both h^4 and h^5. In order for cooperate to hold at both of those nodes, it would have to be that $\rho_2(1-\gamma_2)\phi_{212} > 1 + \rho_2\gamma_2 \cdot 3$ and $\rho_2\gamma_2 \cdot 3 > 1 + \rho_2(1-\gamma_2)\phi_{212}$ which cannot occur. As a result, an equilibrium where player 2 cooperates at both h^4 and h^5 can be ruled out.

If (a) holds in equilibrium, then $x_1 = x_2 = y_1 = y_2 = 1$ and $x_3 = x_4 = y_3 = y_4 = 0$. If player 1 cooperates, then $\phi_{212} = \frac{1}{2}(2(1-2\theta) - 1)$. At h^3 player 2 will cooperate if $3 + (\frac{1}{2})\rho_2(1-\gamma_2)(2(1-2\theta)-1) > 4 - (\frac{1}{2})\rho_2(1-\gamma_2)(2(1-2\theta)-1) - \rho_2 \cdot \gamma_2 \cdot 3$, where $D_{21} = 0$ if player 2 cooperates and $D_{21} = 3$ if player 2 defects. This holds if $\gamma_2 \geq \frac{1-2(1-2\theta))}{4-2(1-2\theta)}$ and $\rho_2 > \frac{1}{2(1-2\theta)-1+\gamma_2(4-2(1-2\theta))}$. At h^4, player 2 will cooperate if $\rho_2(1-\gamma_2)(2(1-2\theta) - 1) - \rho_2 \cdot \gamma_2 \cdot 3 > 1$. This holds if $\gamma_2 \leq \frac{2(1-2\theta)-1}{2(1-2\theta)+2}$, and $\rho_2 > \frac{1}{2(1-2\theta)-1-\gamma_2(2(1-2\theta)+2)}$. Since $\gamma_2 \in [0,1]$, then in order for player 2 to cooperate in pure strategies at h^4 it must be the case that $\theta > \frac{1}{4}$. Since player 2 cooperated at h^4, then it must be the case that player 2 defects at h^5. For player 2 to defect at h^5, then the following must hold $\rho_2(1-\gamma_2)(2(1-2\theta) - 1) - \rho_2 \cdot \gamma_2 \cdot 3 > -1$. Since $\rho_2(1-\gamma_2)(2(1-2\theta)-1) - \rho_2 \cdot \gamma_2 \cdot 3 > 1$, then player 2 will defect at h^5. As a result, if $\theta > \frac{1}{4}$, $0 \leq \gamma_2 \leq \frac{1-4\theta}{4-4\theta}$, and $\rho_2 > \frac{1}{1-4\theta-\gamma_2(4-4\theta)}$, then player 2 will cooperate if player 1 cooperates and defect if player 1 defects.

For (b), in equilibrium it must be the case that $y_1 = 1$ and $y_2 = y_3 = y_4 = 0$. If player 1 cooperates, then $\phi_{212} = \frac{1}{2}(1-3\theta)$. At h^3 player 2 will cooperate if $3 + \rho_2(1-\gamma_2)(1-3\theta) > 4 - \rho_2 \cdot \gamma_2 \cdot 3$. This holds if $\gamma_2 > \frac{3\theta-1}{2+3\theta}$ and $\rho_2 > \frac{1}{1-3\theta+\gamma_2(2+3\theta))}$. At h^4 player 2 will defect if $1 > \rho_2(1-\gamma_2)(1-3\theta) - \rho_2 \cdot \gamma_2 \cdot 3$. This holds if $\gamma_2 < \frac{1-3\theta}{4-3\theta}$ and $\rho_2 < \frac{1}{1-3\theta-\gamma_2(4-3\theta)}$. Since $\gamma_2 \in [0,1]$, in order for player 2 to defect at h^4, then it must be the case that $\theta < \frac{1}{3}$. In order for player 2 to defect at h^5, then it must be the case that $\rho_2(1-\gamma_2)(1-3\theta) - \rho_2 \cdot \gamma_2 \cdot 3 > -1$. This holds if $\rho_2 > \frac{-1}{1-3\theta-\gamma_2(4-3\theta)}$ and $\gamma_2 \leq \frac{1-3\theta}{4-3\theta}$. As a result, if $\theta < \frac{1}{3}$, $0 < \gamma_2 < \frac{1-3\theta}{4-3\theta}$, $\frac{1}{1-3\theta+\gamma_2(2+2\theta)} < \rho_2 < \frac{1}{1-3\theta-\gamma_2(2-4\theta)}$, then player 2 will cooperate if player 1 and nature cooperates and defect otherwise.

For (c), in equilibrium beliefs must be correct, which gives $y_1 = y_3 = 1$ and $y_2 = y_4 = 0$. If player 1 cooperates, then $\phi_{212} = \frac{1}{2}(1-2\theta)$. At h^3 player 2 will cooperate if $\rho_2(1-\gamma_2)(1-2\theta) + \rho_2 \cdot \gamma_2 \cdot 3 > 1$. This holds if $\gamma_2 > \frac{2\theta-1}{2+2\theta}$ and $\rho_2 > \frac{1}{1-2\theta+\gamma_2(2+2\theta)}$. Player 2 will defect at h^4 if $-\rho_2(1-\gamma_2)(1-2\theta) + \rho_2 \cdot \gamma_2 \cdot 3 > -1$. This holds if $\gamma_2 > \frac{1-2\theta}{4-2\theta}$ and $\rho_2 > \frac{-1}{2\theta-1+\gamma_2(4-2\theta)}$. Player 2 will cooperate at h^5 if $-\rho_2(1-\gamma_2)(1-2\theta) + \rho_2 \cdot \gamma_2 \cdot 3 > 1$. This

will hold if $\gamma_2 > \frac{1-2\theta}{4-2\theta}$ and $\rho_2 > \frac{1}{2\theta-1+\gamma_2(4-2\theta)}$. Since $\frac{1}{2\theta-1+\gamma_2(4-2\theta)} \geq \frac{1}{1-2\theta+\gamma_2(2+2\theta)}$, and $\gamma_2 \in [0,1]$, then in order to have this equilibrium it must be the case that$\rho_2 > \frac{1}{2\theta-1+\gamma_2(4+2\theta)}$. So if $\gamma_2 > \frac{1-2\theta}{4-2\theta}$, and t$\rho_2 > \frac{1}{2\theta-1+\gamma_2(4+2\theta)}$, then player 2 will cooperate if nature cooperates and defect if nature defects.

For (d) it must be the case that $y_1 = y_2 = y_3 = y_4 = 0$. If player 1 cooperates, then $\phi_{212} = \frac{1}{2}(2-4\theta)$. At h^3 player 2 will defect if $\rho_2(1-\gamma_2)(2-4\theta) + \rho_2 \cdot \gamma_2 \cdot 3 < 1$. This holds if $\gamma_2 \geq 0$ and $\rho_2 < \frac{1}{2-4\theta+\gamma_2(4\theta-1)}$. At h^4 player 2 will defect if $\rho_2(1-\gamma_2)(2-4\theta))$ $-\rho_2 \cdot \gamma_2 \cdot 3 < 1$ which holds if $\gamma_2 < \frac{2-4\theta}{5-4\theta}$ and $\rho_2 < \frac{1}{2-4\theta-\gamma_2(5-4\theta)}$ or if $\gamma_2 \geq \frac{2-4\theta}{5-4\theta}$ and $\rho_2 \geq 0$. At h^5player 2 will defect if $\rho_2(1-\gamma_2)(2-4\theta) - \rho_2 \cdot \gamma_2 \cdot 3 > -1$. This holds if $\gamma_2 > \frac{2-4\theta}{5-4\theta}$ and $\rho_2 < \frac{1}{4\theta-2+\gamma_2(5-4\theta)}$ or if$\gamma_2 < \frac{2-4\theta}{5-4\theta}$ and $\rho_2 \geq 0$. If $\gamma_2 < \frac{2-4\theta}{5-4\theta}$, then $\frac{1}{2-4\theta-\gamma_2(5-4\theta)} > \frac{1}{2-4\theta+\gamma_2(5-4\theta)}$. If$\gamma_2 > \frac{2-4\theta}{5-4\theta}$, then $\frac{1}{4\theta-2+\gamma_2(5-4\theta)} > \frac{1}{2-4\theta+\gamma_2(4\theta-1)}$. Given this, it follows that if $\gamma_2 > \frac{2-4\theta}{5-4\theta}$, and $\rho_2 < \frac{1}{2-4\theta+\gamma_2(4\theta-1)}$, then player 2 will always defect. If $\gamma_2 < \frac{2-4\theta}{5-4\theta}$ and $\rho_2 < \frac{1}{2-4\theta+\gamma_2(5-4\theta)}$, then player 2 will always defect. □

Concerns for only reciprocity [8] or only inequity aversion [6] arise as special cases. In order to understand the differences between the models of reciprocity and inequity aversion assume that $\gamma_i = 0$. In other words, assume that players are purely reciprocal. As a result of Proposition 1, if player 1 cooperates, then player 2 will cooperate if $\rho_2 > \frac{1}{1-4\theta}$ and $\theta < \frac{1}{4}$. This implies that conditional cooperation by player 2 is only possible provided that player 1's control is sufficiently high. If $\theta > \frac{1}{4}$, then player 2 will not cooperate in pure strategies if player 1 cooperates. For player 2 to interpret a choice by player 1 as kind or unkind, player 1 has to have a certain amount of control over that choice. This model suggests that when control is low, reciprocity is not sufficient to maintain cooperation by player 2. Note that as θ decreases, then ρ_2 must be lower in order to sustain defection as a pure strategy SRE. In other words, as control by player 1 increases, lower concerns for reciprocity are needed for player 2 to always defect.

If players are only inequity averse, then this implies that $\gamma_i = 1$. As result of Proposition 1, inequity aversion predicts that player 2's choice is not influenced by the values of θ. The intended choice of player 1 does not influence what player 2 will choose. Player 2's choice depends only on the degree to which player 2 dislikes getting more than player 1. Cooperation by player 2 is determined by whether player 2 feels "guilty" over receiving more than player 1. If $\rho_2 > \frac{1}{3}$, then player 2 will cooperate if nature cooperates regardless of player 1's choice. That is, the intended choice by player 1 is not behaviorally relevant. This contrasts with the pure reciprocity case in which player 1's intended choice matters for player 2 rather than the results of nature.

If players instead have mixed concerns about reciprocity and inequity aversion, then there are four possible pure strategy equilibria that could hold. If the equilibrium (a) occurs, then player 2 is more concerned about player 1's intentions. This leads to reciprocal behavior where player 2 cooperates if player 1 cooperates and defects if player 1 defects. Provided player 1 has a sufficient level of control over the outcome, this equilibrium is possible. Notice that this equilibrium depends upon player 2's concern for inequity aversion. Lower values of γ_2 suggest that player 2 is more reciprocal; however, if γ_2 is large, then this equilibrium may not occur due to the strong preference for equal outcomes.

The mixed concerns model also suggests another possible equilibrium (b). Here player 2 cooperates if player 1 and nature cooperates, but defects at all other histories. This equilibrium is not possible in the cases of pure reciprocity or pure inequity aversion. In this equilibrium, player 2 cooperates only if player 1 intended to cooperate and the result of that intention leads to cooperation. Intentions are not enough for player 2 to cooperate when player 1 cooperates and nature defects. In addition, if the concern about inequity aversion is sufficiently small, then player 2 will not cooperate if player 1 defects and nature cooperates. Here player 2 may be concerned about both intentions and the distribution of outcomes, but cooperation is only sustained when those concerns align.

The equilibrium (c) occurs if players are strongly inequity aversion averse. One thing to notice is that this equilibrium has no restrictions on the value of θ other than the assumption that $\theta > \frac{1}{2}$. Since the mixed concerns model allows inequity aversion and reciprocity, player 2 must have a sufficiently high concern for inequity aversion in order for (3) to hold. One interesting result is that as the value of θ increases, this equilibrium holds for smaller values of γ_2. This result makes intuitive sense. To see why, suppose that player 2 is really concerned about reciprocity. When player 1 has little control, player 1's choice is not seen as very intentional. Consequently, reciprocity has little weight in player 2's decision. As a result, inequity aversion can become more important as first mover control decreases. Since reciprocity is not much of a factor when control is low, concerns for reciprocity do not conflict as much with concerns for inequity aversion at nodes h^4 and h^5.

The equilibrium (d) gives the case when player 2 will always defect. If $\rho_2 = 0$, then the model is just the self-interest model and player 2 will always defect. If $\rho_2 > 0$, then the minimum value of ρ_2 that will lead to player 2 always defecting depends on the relative weight they place on the two concerns and the reversal probability.

Appendix A.3.2. Imperfect Information

In the imperfect information game, player 2 does not know what player 1 chose but does know the results of nature. In any SRE, the potential behavior for player 2 is described in Proposition 2.

Proposition A5. *If $\theta < \frac{1}{2}$, then in any SRE the potential behavior for player 2 can be described as follows:*

1. *If player 1 cooperates*
   (a) *If $0 < \gamma_2 < \frac{1-2\theta}{4-2\theta}$ and $\frac{1}{1-2\theta+\gamma_2(2+2\theta)} < \rho_2 < \frac{1}{1-2\theta-\gamma_2(4-2\theta)}$ or $\gamma_2 > \frac{1-2\theta}{4-2\theta}$ and $\rho_2 > \frac{1}{1-2\theta+\gamma_2(2+2\theta)}$, then player 2 will cooperate if nature cooperates and defect if nature defects.*
   (b) *If $\gamma_2 < \frac{2-4\theta}{5-4\theta}$ and $\rho_2 > \frac{1}{2-4\theta-\gamma_2(5-4\theta)}$, then player 2 will always cooperate.*
   (c) If $\gamma_2 \geq 0$ and $\rho_2 < \frac{1}{2-4\theta+\gamma(1+4\theta)}$, *then player 2 will always defect.*
2. *If player 1 defects*
   (a) If $\gamma_2 > \frac{1-2\theta}{4-2\theta}$ *and* $\rho_2 > \frac{1}{\gamma_2(4-2\theta)-1+2\theta}$, *then player 2 will cooperate if nature cooperates and defect if nature defects.*
   (b) If $\gamma_2 > \frac{2-4\theta}{5-4\theta}$ *and* $\rho_2 < \frac{1}{\gamma_2(5-4\theta)-2+4\theta}$ *or* $\gamma_2 < \frac{2-4\theta}{5-4\theta}$ *and* $\rho_2 \geq 0$, *then player 2 will always defect.*

Proof. Let player 1's belief about what player 2 will choose at each information set be defined as: $q_1 = P_1(2\,choses\,C|h^3 \cup h^5)$, and $q_2 = P_1(2\,choses\,C|h^4 \cup h^6)$. Player 2's belief about player 1's belief about what player 2 will choose at each information set is defined as the expectation of player 1's beliefs about player 2. This gives: $v_1 = E_2[q_1|h^3 \cup h^5]$, and $v_2 = E_2[q_2|h^4 \cup h^6]$. Player 1 can choose to cooperate or defect at node h^0. Let player 2's belief that player 1 will cooperate be $z_1 = P_2(1\,choses\,C|h^0)$. Player 1's belief about player 2's belief that player 1 will cooperate is defined as $w_1 = E_1[z_1|h^0]$.

Player 2 only observes the results of nature. Player 2's evaluation of the kindness of player 1 depends on the belief about what node she is currently at. If player 2 observes cooperation, then the probability that player 2 believes she is at node h^3 is $P(h^3|2\,observes\,C) = \frac{z_1 \cdot \varepsilon_1}{z_1 \cdot \varepsilon_1 + (1-z_1) \cdot \varepsilon_2}$ via Bayes rule. Similarly, $P(h^5|2\,observes\,C) = \frac{(1-z_1) \cdot \varepsilon_2}{z_1 \cdot \varepsilon_1 + (1-z_1) \cdot \varepsilon_2}$. If player 2 observes defection, then $P(h^4|2\,observes\,D) = \frac{z_1 \cdot (1-\varepsilon_1)}{z_1 \cdot (1-\varepsilon_1) + (1-z_1) \cdot (1-\varepsilon_2)}$ and $P(h^6|2\,observes\,D) = \frac{(1-z_1) \cdot (1-\varepsilon_2)}{z_1 \cdot (1-\varepsilon_1) + (1-z_1) \cdot (1-\varepsilon_2)}$. Since the SRE concept requires that initial beliefs be correct, it follows that player 2 knows in equilibrium what player 1 chooses.

Player 2's belief about the kindness of player 1 is $\phi_{212} = (z_1 - \frac{1}{2})(1-2\theta)(2 - v_1 + v_2)$. No matter the history, the kindness of player 2 towards player 1 will always be $k_{21} = 1$ if player 2 cooperates and $k_{21} = -1$ if player 2 defects. If at nodes h^3 and h^5, then $D_{21} = -3$ if player 2 defects and zero otherwise. If at nodes h^4 and h^6, then if player 1 cooperates $D_{21} = -3$ and is equal to zero otherwise.

Suppose that in equilibrium player 1 cooperates, then player 2 knows this. Since $z_1 = w_1 = 1$, player 2 knows that if nature cooperates then she is at node h^3 and if nature defects then she is at node h^4. If nature cooperates, then player 2 will cooperate if $(1-2\theta)(2 - v_1 + v_2)\rho_2(1-\gamma_2) + 3\gamma_2\rho_2 > 1$. If nature defects, then player 2 will cooperate if $(1-2\theta)(2 - v_1 + v_2)\rho_2(1-\gamma_2) > 1 + 3\gamma_2\rho_2$.

For 1(a), $v_1 = 1$ and $v_2 = 0$. For this to be an equilibrium it must be that $(1-2\theta)\rho_2(1-\gamma_2)$ $+3\gamma_2\rho_2 > 1$ and $(1-2\theta)\rho_2(1-\gamma_2) < 1 + 3\gamma_2\rho_2$. This hold under two conditions. In the first case, if $0 < \gamma_2 < \frac{1-2\theta}{4-2\theta}$, and $\frac{1}{1-2\theta+\gamma_2(2+2\theta)} < \rho_2 < \frac{1}{1-2\theta-\gamma_2(4-2\theta)}$, then player 2 will cooperate if nature cooperates and defect if nature defects. In the second case, if$\gamma_2 > \frac{1-2\theta}{4-2\theta}$ and $\rho_2 > \frac{1}{1-2\theta+\gamma_2(2+2\theta)}$, then player 2 will cooperate if nature cooperates and defect if nature defects.

For 1(b), $v_1 = v_2 = 1$. This is possible if $(1-2\theta)(2)\rho_2(1-\gamma_2) + 3\gamma_2\rho_2 > 1$ and $(1-2\theta)(2)\rho_2(1-\gamma_2) > 1 + 3\gamma_2\rho_2$. If $\gamma_2 < \frac{2-4\theta}{5-4\theta}$ and$\rho_2 > \frac{1}{2-4\theta-\gamma_2(5-4\theta)}$, then both conditions will be satisfied. Player 1 will cooperate no matter the results of nature.

For 1(c), $v_1 = v_2 = 0$. This is possible if both $(1-2\theta)(2)\rho_2(1-\gamma_2) + 3\gamma_2\rho_2 < 1$ and $(1-2\theta)(2)\rho_2(1-\gamma_2) < 1 + 3\gamma_2\rho_2$. If $\gamma_2 \geq 0$ and $\rho_2 < \frac{1}{2-4\theta-\gamma_2(1-4\theta)}$, then player 2 will always defect.

If player 1 defects, then player 2's belief about the kindness of player 1 is$\phi_{212} = -\frac{1}{2}(1-2\theta)(2 - v_1 + v_2)$. For 2(a), $v_1 = 1$ and $v_2 = 0$. This implies that $-(1-2\theta)\rho_2(1-\gamma_2)$ $+3\gamma_2\rho_2 > 1$ and $(1-2\theta)(2)\rho_2(1-\gamma_2) + 1 + 3\gamma_2\rho_2 > 0$. These conditions will hold if $\gamma_2 > \frac{1-2\theta}{4-2\theta}$ and $\rho_2 > \frac{1}{\gamma_2(4-2\theta)-1+2\theta}$.

For 2(b), $v_1 = v_2 = 0$. This is possible if both $-(1-2\theta)(2)\rho_2(1-\gamma_2) + 3\gamma_2\rho_2 < 1$ and $-(1-2\theta)(2)\rho_2(1-\gamma_2) < 1 + 3\gamma_2\rho_2$. These conditions will hold if$\gamma_2 > \frac{2-4\theta}{5-4\theta}$ and $\rho_2 < \frac{1}{\gamma_2(5-4\theta)-2+4\theta}$ or $\gamma_2 < \frac{2-4\theta}{5-4\theta}$ and$\rho_2 \geq 0$. □

To understand the equilibrium predictions when players are purely reciprocal, assume that $\gamma_i = 0$. With pure reciprocity, player 2 ignores the results of nature. As a consequence, player 2 will choose to cooperate based on the equilibrium beliefs about what player 1 chose. If player 1 cooperates with probability one, then player 1 is being kind towards player 2. Even if player 2 observes defection by nature, player 2 knows that player 1 cooperated and player 1 is still viewed as kind.

The control that player 1 has still matters. When player 1 has more control, the value of ρ_2 needed for player 2 to cooperate can be smaller all other things equal. This suggests that cooperation should be higher when player 1 has more control. Notice however that cooperation in pure strategies is still possible even when control is low. This differs from the perfect information game.

If players are purely inequity averse, then $\gamma_i = 1$. The equilibrium predictions for a player 2 with pure inequity aversion are the same for the perfect or imperfect information games. This makes sense because inequity aversion is only outcome based, and player 1's intended choice does not influence player 2's fairness judgments.

With mixed concerns, the potential equilibrium in 1(a) gives that player 2 will cooperate if nature cooperates and defect if nature defects. This equilibrium can occur if player 2 is strongly concerned about inequity aversion. Notice, however, that the equilibrium is also possible for a player 2 that cares a great deal about reciprocity. For certain ranges of ρ_2, a player that is highly reciprocal will behave as if they are concerned about inequity aversion. This suggests that as control changes the types that players appear to be could change as well. As a result, it is possible that some players could behave inequity averse, self-interested, or reciprocal depending upon player 1's level of control. In 1(b), player 2 will cooperate regardless of nature's choice. In equilibrium, player 2 knows that player 1

cooperated and cooperation by player 1 is viewed as kind. This kindness is enough for players that are highly concerned about reciprocity to cooperate even if nature defects.

There are a large number of potential equilibria that can occur for player 1 due to self-fulfilling expectations. The focus of this paper on second mover behavior. In the interest of space, equilibrium predictions for player 1 are available upon request.

Appendix A.3.3. Perfect Versus Imperfect Information

Both the perfect and imperfect information games can be used to test the predictions of the fairness models explored in this paper. Predictions from pure self-interest and pure inequity aversion are the same no matter the information. As a result of these models, changes in the information about what player 1 chose should not be relevant for equilibrium behavior.

With pure reciprocity, equilibrium behavior could differ depending upon the information available to player 2. In the perfect information game, pure strategy cooperation by player 2 only occurs if control is high. However, in the imperfect information game, cooperation is still possible when control is low. Even when control is high, the concern for reciprocity needs to be much higher when information is perfect compared to the imperfect information game in order for cooperation to be possible. As a result, if subjects are motivated by reciprocity, then cooperation should be higher when information is imperfect compared to when the information is perfect.

In the mixed concerns model, when control is high in the perfect information game, it is possible to have an equilibrium in which player 2 cooperates if player 1 cooperates and defects if player 1 defects. However, when control is low this equilibrium no longer exists. This is not the case with the imperfect information game. When control is low it is still possible for player 2 to cooperate if player 1 cooperates and defect if player 1 defects. Even when control is high, the range of values for both ρ_2 and γ_2 that lead player 2 to cooperate is largest in the imperfect information game. Thus, given that player 1 cooperates, cooperation by player 2 in the imperfect information game should be higher than in the perfect information game.

Notes

1 Examples of outcome-based models include Bolton2000 and Fehr and Schmidt [6], and intention-based models include Rabin [7], and Dufwenberg and Kirchsteiger [8]. Models that combine concerns for intentions and outcomes include Levine [9], Charness and Rabin [10], Falk and Fischbacher [11], and Cox et al. [12].

2 While this paper assumes the reversal probability is the same whether the first mover cooperates or defects. Theoretical results are similar if the reversal probabilities are allowed to differ. For clarity of presentation, a single probability θ is used both in the theoretical analysis and the experiment.

3 The analysis can be done without this restriction. However, if $\theta > \frac{1}{2}$, then the choice that the first mover chooses is more likely to be switched. While this makes sense mathematically, it is not clear that this represents what occurs in most human interactions. Having $\theta > \frac{1}{2}$ means that if the first mover wants nature to be more likely to choose cooperate, then the first mover should defect. This is not say these types of situations cannot occur, but the main focus of the paper will be when a player's intended choice matches the player's actual choice.

4 The reversal probability of 10% corresponds to Figures 1 and 2 where $\theta = 0.1$.

5 In addition, these results are robust to including age, number of economic classes, number of statistics classes, and political views.

6 Here the only difference between a_i and $a_i(h)$ is that choices in history h are made with probability one.

7 Here the equitable payoff is mathematically equivalent to (3).

8 Many different types of distributional concerns could be considered. Other forms to be included could be Rawlsian, Utilitarian, or Nash Product.

9 Assuming the standard function form for inequity aversion Fehr and Schmidt [6] gives the same equilibrium predictions for second movers in the sequential prisoner's dilemma with nature as the restricted functional form assumed here.

References

1. Ahn, T.; Lee, M.; Ruttan, L.; Walker, J. Asymmetric payoffs in simultaneous and sequential prisoner's dilemma games. *Public Choice* **2007**, *132*, 353–366. [CrossRef]
2. Bolle, F.; Ockenfels, P. Prisoners' dilemma as a game with incomplete information. *J. Econ. Psychol.* **1990**, *11*, 69–84. [CrossRef]
3. Clark, K.; Sefton, M. The sequential prisoner's dilemma: Evidence on reciprocation. *Econ. J.* **2001**, *111*, 51–68. [CrossRef]
4. Ridinger, G.; McBride, M. Reciprocity in Games with Unknown Types. In *Handbook of Experimental Game Theory*; Capra, M., Croson, R., Rigdon, M., Rosenblatt, T., Eds.; Edward Elgar Publishing: Cheltenham, UK; Northampton, MA, USA, 2020.
5. Bolton, G.E.; Okenfels, A. Erc: A theory of equity, reciprocity, and competition. *Am. Econ. Rev.* **2000**, *90*, 166–193. [CrossRef]
6. Fehr, E.; Schmidt, K.M. A theory of fairness, competition, and cooperation. *Q. J. Econ.* **1999**, *114*, 817–868. [CrossRef]
7. Rabin, M. Incorporating fairness into game theory and economics. *Am. Econ. Rev.* **1993**, *83*, 1281–1302.
8. Dufwenberg, M.; Kirchsteiger, G. A theory of sequential reciprocity. *Games Econ. Behav.* **2004**, *47*, 268–298. [CrossRef]
9. Levine, D.K. Modeling altruism and spitefulness in experiments. *Rev. Econ. Dyn.* **1998**, *1*, 593–622. [CrossRef]
10. Charness, G.; Rabin, M. Understanding social preferences with simple tests. *Q. J. Econ.* **2002**, *117*, 817–869. [CrossRef]
11. Falk, A.; Fischbacher, U. A theory of reciprocity. *Games Econ. Behav.* **2006**, *54*, 293–315. [CrossRef]
12. Cox, J.C.; Friedman, D.; Gjerstad, S. A tractable model of reciprocity and fairness. *Games Econ. Behav.* **2007**, *59*, 17–45. [CrossRef]
13. Chaudhuri, A. Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Exp. Econ.* **2011**, *14*, 47–83. [CrossRef]
14. Chaudhuri, A.; Paichayontvijit, T. Conditional cooperation and voluntary contributions to a public good. *Econ. Bull.* **2006**, *3*, 1–14.
15. Fischbacher, U.; Gachter, S.; Fehr, E. Are people conditionally cooperative? evidence from a public goods experiment. *Econ. Lett.* **2001**, *71*, 397–404. [CrossRef]
16. Herrmann, B.; Thoni, C. Measuring conditional cooperation: A replication study in russia. *Exp. Econ.* **2009**, *12*, 87–92. [CrossRef]
17. Rustagi, D.; Engel, S.; Kosfeld, M. Conditional coopeation and cocost monitorying explain success in forest ccommon management. *Science* **2010**, *330*, 961–965. [CrossRef] [PubMed]
18. Dhaene, G.; Bouckaert, J. Sequential reciprocity in two-player, two-stage games: An experimental analysis. *Games Econ. Behav.* **2010**, *70*, 289–303. [CrossRef]
19. Blanco, M.; Engelmann, D.; Normann, H.T. A within-subject analysis of other-regarding preferences. *Games Econ. Behav.* **2011**, *72*, 321–338. [CrossRef]
20. Falk, A.; Fehr, E.; Fischbacher, U. On the nature of fair behavior. *Econ. Inq.* **2003**, *41*, 20–26. [CrossRef]
21. Bolton, G.E.; Okenfels, A. A stress test of fairness measures in models of social utility. *Econ. Theory* **2005**, *25*, 2005. [CrossRef]
22. Falk, A.; Kosfeld, M. The hidden costs of control. *Am. Econ. Rev.* **2006**, *96*, 1611–1630. [CrossRef]
23. Stanca, L. How to be kind? Outcomes versus intentions as determinants of fairness. *Econ. Lett.* **2010**, *106*, 19–21. [CrossRef]
24. Charness, G. Attribution and reciprocity in an experimental labor market. *J. Labor Econ.* **2004**, *22*, 665–688. [CrossRef]
25. Falk, A.; Fehr, E.; Fischbacher, U. Testing theories of fairness- intentions matter. *Games Econ. Behav.* **2008**, *62*, 287–303. [CrossRef]
26. Bolton, G.E.; Brandts, J.; Okenfels, A. Measuring motivations for the reciprocal responses observed in a simple dilemma game. *Exp. Econ.* **1998**, *1*, 207–219. [CrossRef]
27. Charness, G.; Levine, D.I. Intention and stochastic outcomes: An experimental study. *Econ. J.* **2007**, *117*, 1051–1072. [CrossRef]
28. Cox, J.C.; Deck, C.A. Assigning intentions when actions are unobservable: The impact of trembling in the trust game. *South. Econ. J.* **2006**, *73*, 307–314.
29. Blanco, M.; Engelmann, D.; Koch, A.K.; Normann, H.T. Preferences and beliefs in a sequential social dilemma: A within-subject analysis. *Games Econ. Behav.* **2001**, *87*, 122–135. [CrossRef]
30. Ridinger, G.; McBride, M. Theory of Mind Ability and Cooperation. Working Paper. 2017. Available online: https://economics.ucr.edu/wp-content/uploads/2019/10/McBride-paper-for-1-31-18-seminar.pdf (accessed on 10 January 2020).
31. Sher, I.; Koenig, M.; Rustichini, A. Children's strategic theory of mind. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 13307–13312. [CrossRef]
32. Batson, D. These things called empathy: Eight related but distinct phenomenon. In *The Social Neuroscience of Empathy*; Decety, J., Ickes, W., Eds.; MIT Press: Cambridge, UK, 2011.
33. Ridinger, G.; McBride, M. Money Affects Theory of Mind Differently by Gender. *PLoS ONE* **2015**, *10*, e0143973. [CrossRef]
34. Ridinger, G. Shame and Theory-of-Mind Predicts Rule-following Behavior. *Games* **2020**, *11*, 36. [CrossRef]
35. Singer, T. Neuroeconomics: Decision Making and the Brain. In *Understanding Others: Brain Mechanisms of Theory of Mind and Empathy*; Academic Press: London, UK, 2009; pp. 251–265.
36. Smith, A. The Theory of Moral Sentiments. London: A. Millar. Library of Economics and Liberty. 1790. Available online: http://www.econlib.org/library/Smith/smMS.html (accessed on 24 September 2014)
37. Singer, T.; Steinbeis, N. Differential roles of fairness- and compassion-based motivations for cooperation, defection, and punishment. *Ann. N. Y. Acad. Sci.* **2009**, *1167*, 41–50. [CrossRef]
38. Batson, D.C.; Eklund, J.H.; Chermok, V.L.; Hoyt, J.L.; Ortiz, B.G. An additional antecedent of empathic concern: Valuing the welfare of the person in need. *J. Personal. Soc. Psychol.* **2007**, *93*, 65–74. [CrossRef]
39. Bekkers, R. Traditional and health-related philanthropy: The role of resources and personality. *Soc. Psychol. Q.* **2006**, *69*, 349–366. [CrossRef]

40. Ridinger, G. Empathetic Concern, Altruism, and the Pursuit of Distributive Justice. Master's Thesis, California State University, Fullerton, CA, USA, 2011.
41. Kruger, D.J. Evolution and altruism: Combining psychological mediators with naturally selected tendences. *Evol. Hum. Behav.* **2003**, *24*, 118–125. [CrossRef]
42. Bartels, D.M.; Kvaran, T.; Nichols, S. Selfless giving. *Cognition* **2013**, *129*, 392–403. [CrossRef]
43. Leliveld, M.C.; Vandijk, E.; Vanbeest, I. Punishing and compensating others at your own expense: The role of empathic concern on reactions to distributive justice. *Eur. J. Soc. Psychol.* **2012**, *42*, 135–140. [CrossRef]
44. Batson, D.C.; Batson, J.G.; Todd, M.R.; Brummett, B.H.; Shaw, L.L.; Aldeguer, C.M.R. Empathy and the collective good: Caring for one of the others in a social dilemma. *J. Personal. Soc. Psychol.* **1995**, *68*, 619–631. [CrossRef]
45. Oceja, L.; Jimenez, I. Beyond egoism and gruop identity: Empathy toward the other and awareness of others in a social dilemma. *Span. J. Psychol.* **2007**, *10*, 369–379. [CrossRef] [PubMed]
46. Batson, D.C.; Moran, T. Empathy-induced altruism in a prisoner's dilemma. *Eur. J. Soc. Psychol.* **1999**, *22*, 474–482. [CrossRef]
47. Rumble, A.C.; Lange, P.A.M.V.; Parks, C.D. The benefits of empathy: When empathy may sustain cooperation in social dilemmas. *Eur. J. Soc. Psychol.* **2010**, *40*, 856–866. [CrossRef]
48. Batson, D.; Ahmad, N. Empathy-induced altruism in a prisoner's dilemma ii: What if the target of empathy has defected? *Eur. J. Soc. Psychol.* **2001**, *31*, 25–36. [CrossRef]
49. Fischbacher, U. z-tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* **2007**, *10*, 171–178. [CrossRef]
50. Davis, M.H. Measuring individual differences in empathy: Evidence for a multidimensional approach. *J. Personal. Soc. Psychol.* **1983**, *44*, 113–126. [CrossRef]
51. Bicchieri, C. *The Grammar of Society: The Nature and Dynamics of Social Norms*; Cambridge University Press: Cambridge, UK, 2006.
52. Ridinger, G. Ownership, punishment, and norms in a real-effort bargaining experiment. *J. Econ. Behav.* **2018**, *155*, 382–402. [CrossRef]
53. Rand, D.G.; Greene, J.D.; Nowak, M.A. Spontaneous giving and calculated greed. *Nature* **2005**, *489*, 427–430. [CrossRef]
54. Dana, J.; Weber, R.A.; Kuang, J.X. Exploting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Econ. Theory* **2007**, *33*, 67–80. [CrossRef]
55. Camerer, C.F. *Behavioral Game Theory: Experiments in Strategic Interaction*; Princeton University Press: Princeton, NJ, USA, 2003.
56. Battigalli, P.; Dufwenberg, M. Dynamic psychological games. *J. Econ. Theory* **2009**, *144*, 1–35. [CrossRef]
57. Geanakoplos, J.; Pearce, D.; Stacchetti, E. Psychological games and sequential rationality. *Games Econ. Behav.* **1989**, *1*, 60–79. [CrossRef]
58. Sebald, A. Attribution and reciprocity. *Games Econ. Behav.* **2010**, *68*, 339–352. [CrossRef]