Essays in behavioral economics

Mermer, A.G.

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
Mermer, A. G. (2014). *Essays in behavioral economics: Applied game theory and experiments*. CentER, Center for Economic Research.

Essays in Behavioral Economics: Applied Game Theory and Experiments

Ayşe Gül Mermer

December 18, 2014

Essays in Behavioral Economics: Applied Game Theory and Experiments

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof.dr. Ph. Eijlander, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op donderdag 18 december 2014 om 10.15 uur door

Ayşe Gül Mermer

geboren op 10 mei 1983 te Konya, Turkije.

Promotiecommissie:

Promotor:	prof.dr. W. Müller
Copromotor:	dr. S. Suetens
Overige Leden:	prof.dr. J. Boone
	prof.dr. D. Engelmann
	prof.dr. C.N. Noussair
	prof.dr. J.J.M. Potters
	prof.dr. R. Sloof

Acknowledgements

First of all, I would like to express my gratitude towards my advisors Wieland Müller and Sigrid Suetens for their excellent guidance and continuous support.

Wieland, your dedication to and enthusiasm for research about research immensely impressed me always and was an important factor motivating me towards my academic life. In the majority of my studies, I did not have the chance to work with you under the same roof, which I believe is a big loss for me. However, you taught me that geographical distance does not matter when it comes to research. You were always very generous in making time available and calling me, even on very short notice, to discuss matters whenever I needed advice. You patiently taught me how to find my way when I was lost. I am very thankful to you for inspiring me to work on my job market paper yet giving me the freedom to develop my own paper. I am very lucky to have the chance of co-authoring a paper with you, where I learned the way to think about economic problems and to pay great attention to details.

Sigrid, I greatly admire your conscientious attitude towards research and very much appreciate having you as a successful female role model for me. It is hard to find words that would describe your tireless effort and your dedication to my academic development. I always feel privileged that I could co-author papers with you, through which I learned how to work on a research project. I not only learned a lot from you, but also enjoyed working with you a lot. I very much hope that we will continue to work together in the future. You always made time available to discuss my research and academic life, including weekends and late hours whenever I was in need of advice. I am always impressed by your constantly positive attitude towards me; believing in me even in my desperate times. You always gave me freedom to develop my own way in research, but also stepped in to help whenever I needed it.

I highly appreciate the strong support of Wieland and Sigrid when it came to summer schools, conferences, seminars and the stressful job market period.

My gratitude goes also to the committee members, Jan Boone, Dirk Engelmann, Charles Noussair, Jan Potters, Randolph Sloof for their time and efforts in reviewing my manuscript. Charles, I always admire your enthusiasm towards research and how you enjoy discussing any topic related to economics. I always felt enlightened and very motivated after discussing my papers with you. I also would like to acknowledge your invaluable advice on my PhD and the job market in the role of director of graduate studies and I am very thankful for your strong support during the job market. Jan (Boone), I always felt lucky when I had the opportunity to discuss my research with you. I am thankful for all your valuable suggestions and comments on my papers and for your support in the job market. Jan (Potters), you played an important role towards my PhD by introducing the world of behavioral and experimental economics to me. It was always a pleasure to listen to your lectures. I am very thankful for all your suggestions on the drafts of my papers and on my experimental designs. I would like to thank Randolph for the very helpful discussion and comments during my pre-defense and to Dirk for his detailed feedback and very helpful suggestions on my manuscript.

I benefitted from comments by many other people. I would like to thank Cedric Argenton, Johannes Binswanger, Patricio Dalton, Erik van Damme, Sebastian Ebert, Gijs van de Kuilen, Peter Kooreman, Jens Prüfer, Florian Schütt, Stefan Trautman, and Bert Williams for their helpful comments on my papers. I am thankful to the Tilburg PhD Job Market Committee: Charles Noussair, Patricio Dalton, Otilia Boldea, Burak Uras and Damjan Pfajfar for their support and useful advice. I also thank Meltem Daysal for giving me invaluable tips on the job market and her support. I would like to thank all my fellow PhD students in Tilburg for making my PhD life much easier and a lot more fun. I also would like to thank the administrative staff of the Economics Department and CentER for their help during my PhD.

Last but not least, my deep gratitude goes to my parents and my sister for always being there for me, for their endless support and love. Canim anneciğim ve babaciğim, canımın içi Bilge'm, sonsuz sabrınız, sevginiz ve desteğiniz için çok teşekkür ederim. And finally, I express my deep gratitude to Christoph Schottmüller for always being there for me, for his invaluable suggestions on my papers, for his endless support and understanding in the last three years and for giving me the best memories of Tilburg.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Traditional economic theory builds on models of rational decision makers who maximize their monetary utilities without having any other concerns. The validity of the predictions of these models has been tested since the early 1930s using laboratory experiments. It has been well established that the observed behavior in the lab is not always in line with the predictions of standard economic models. This led economists to question the underlying behavioral assumptions.

Behavioral Economics aims at understanding the decisions of economic agents who are not necessarily monetary utility maximizers and accounts for the fact that agents may have other concerns in addition to economic gain. It integrates insights from other fields studying human behavior into economics. The current thesis consists of three chapters that aim at understanding the decisions of economics agents who are not necessarily monetary utility maximizers in situations with strategic interaction.

A first method used by behavioral economists is to develop theoretical models that use non-standard preferences that have been found to align empirical evidence. Chapter 2 of this thesis relates to this point and solves a game-theoretic model assuming that agents have reference dependent preferences. The results help to explain behavior observed in various experiments that is hard to reconcile with the assumption of standard preferences.

A second method used by behavioral economists is laboratory experimentation which allows for careful scrutinizing of behavioral assumptions made in economic models. Chapter 3 and 4 fit within this line of research. In Chapter 3 we experimentally investigate agents' behavior in dilemma games with different strategic environments. In Chapter 4 we experimentally study information acquisition in a social dilemma game. In what follows each chapter is summarized in turn.

Chapter 2 (single-authored) studies a multiple prize contest assuming that agents have expectation-based reference-dependent preferences a la Koszegi and Rabin (2006). In a contest game, first the contest designer decides on the prize structure - the number and the level of prizes -, and then contestants simultaneously undertake costly efforts. Each contestant has private information about his ability which affects his cost-of-effort. The model provides an explanation for the observed behavior in recent laboratory experiments. In particular, that high-ability contestants overexert effort while low-ability contestants exert very little or no effort in comparison to predictions with standard preferences. I also show that the optimal prize allocation in contests may differ markedly in the presence of expectation-based loss aversion. In particular, I show that multiple prizes can be optimal when the cost-of-effort function is linear or concave, where standard preferences predict the optimality of a single prize in these cases. Several unequal prizes might be optimal when the cost-of-effort function is convex.

Chapter 3 (co-authored with Wieland Muller and Sigrid Suetens) uses a laboratory experiment to study the effect of strategic substitutability and strategic complementarity on the extent of cooperative behavior in indefinitely repeated two-player games. On average, choices in our experiment do not differ between the strategic complements and substitutes treatments. However, the aggregate data mask two countervailing effects. On the one hand, the percentage of fully cooperative choices is significantly higher under strategic substitutes than under strategic complements. We argue that this difference is driven by the fact that it is less risky to cooperate under substitutes than under complements. On the other hand, choices of subjects in pairs that do not succeed in cooperating at the joint-payoff maximum tend to be lower, i.e. less cooperative, under strategic substitutes than under strategic complements. We relate the latter result to non-equilibrium forces stemming from a combination of heterogeneity of subjects and differences in the slope of the best-response function between substitutes and complements.

Chapter 4 (co-authored with Sigrid Suetens) uses laboratory experiments to study the behavior of agents in a trust game. We design an experiment to study whether trustors choose to be informed about the type of the trustee in a twice-repeated trust game where, theoretically, having such information is detrimental for cooperation and material payoffs. In one treatment trustors are not informed about the type of the matched trustee and in another treatment they have the choice to obtain information about the type. We find that almost all subjects in the role of trustors choose to obtain the information if they have the chance to do so. We also find that trustors who are informed about the type trust less than the ones who are uninformed.

Chapter 2

Contests with Expectation-Based Loss-Averse Players

2.1. Introduction

A contest is an event where participants compete with each other by means of exerting costly efforts in order to win prizes. There are many economic and social environments that could be described as contests. In sports, athletes compete with each other for gold, silver and bronze medals, and in firms, employees exert effort in order to be promoted to certain positions. In these examples, the contest designer's motive in choosing the prize structure is to increase contestants' performance, for example, to thrill the audience in sports contests or to obtain the highest output in firms. Since such competitive environments are prevalent in many contexts, contests and their design are studied extensively in the economic literature both theoretically and experimentally.

An important common finding in several experimental studies is the discrepancy between behavior predicted by theory and behavior observed in the lab. In particular, high-ability subjects spend more effort while low-ability subjects spend less or no effort in comparison to predictions with standard preferences (e.g. Barut and Noussair 2002, Noussair and Silver 2006, Ernst and Thoni 2009, Müller and Schotter 2010, Klose and Sheremeta 2012, Schram and Ondersal 2009). Some of these studies suggest that this discrepancy may be caused by loss aversion on the part of subjects. One prominent model of loss aversion is Kőszegi and Rabin's model of reference dependent preferences. In this model, next to the standard consumption utility, the agent derives gain-loss utility by comparing outcomes to his reference point. A key assumption of this model is that agent's reference point is his rational expectations.

Recent empirical studies provide evidence for expectations being determinants of agent's reference point. Post et al. (2008) examine the behavior of contestants in the TV show “Deal or No Deal”. They find that contestants' choices can be explained largely by experienced previous outcomes. Their result suggests that lagged expectations serve as a reference point for contestants as predicted by expectation-based reference-dependent preferences. Abeler et al. (2011) conduct a real-effort experiment to test whether a change in expectations of subjects affect their effort provision. To do so, they manipulate the rational expectations of subjects. They find that subjects with high expectations work harder and longer than subjects with low expectations, in line with the predictions of expectation-based reference-dependent preferences.

In this paper, I generalize Moldovanu and Sela's (2001) contest model, by allowing for expectation-based loss aversion á la Kőszegi and Rabin (2006) on the part of the contestants. My model predicts that high-ability contestants exert more effort, while low-ability contestants exert very little or no effort relative to the predictions with standard preferences. This result is consistent with the behavior observed in recent laboratory experiments. The effort provision of the contestants has important implications for the optimal design of the prize structure. In fact, I show that the optimal allocation of prizes in a contest changes markedly when contestants are expectation-based loss-averse. In particular, multiple prizes can be optimal when the cost-of-effort function is either linear or concave, where standard preferences predict the optimality of a single prize.

Moldovanu and Sela (2001) (henceforth M-S) consider the following contest model. The contest designer first determines the allocation of prizes (the number and the level of prizes) given a fixed total prize sum. The goal of the contest designer is to maximize the total expected effort of the contestants. Given the prize structure, contestants with standard preferences choose their effort level in order to maximize their expected utility. The contestant with the highest effort wins first prize, the contestant with the second highest effort wins second prize, and so on until all prizes are distributed. Each contestant bears the cost of effort regardless of winning a prize or not. The cost-of-effort function depends on the ability parameter, which is private information, as well as the effort level. In this model, I introduce expectation-based loss aversion on the part of contestants in the sense of Kőszegi and Rabin (2006) (henceforth K-R). Following K-R, each contestant, next to the standard consumption utility, derives a gain-loss utility by comparing the actual outcome with his expectations. More specifically, each contestant compares the realized outcome with all other possible outcomes that could have occurred and weights each of these comparisons with the *ex-ante* probability of the alternative outcome occurring. Incorporating expectations as the agent's reference point induces a bifurcating force among the efforts of high- and low-ability contestants. Intuitively, a high-ability contestant, who has an *ex-ante* high chance of winning a prize, holds high expectations for winning a prize. In order to avoid the loss sensation associated with not winning a prize, he increases his effort level to further increase his chances of winning. A low-ability contestant, who has an *ex-ante* low chance of winning a prize, holds low expectations for winning a prize. In order to avoid the feeling of losing a prize, he decreases his effort level to further decrease his expectations. Moreover, if a low-ability contestant is sufficiently loss-averse, the gain-loss utility might dominate the consumption utility. In this case, a contestant exerting positive effort might end up with a negative expected utility. In order to avoid this, he reduces his effort level to the minimum possible level and exerts zero effort.[1]

The contest designer, anticipating the contestants' behavior, aims to maximize the total expected effort of the contestants. Thus, any change in the contestants' effort provision has important implications on the designer's decision about prize allocation. I show that, in the presence of expectation-based loss aversion, multiple prizes can be optimal when the cost-of-effort functions are linear or concave, whereas, with standard preferences, a single prize is optimal in these cases. Intuitively, if a single prize is announced by the designer, a low-ability contestant loses the slim hope of winning the prize and exerts very little or no effort. However, a high-ability contestant exerts effort aggressively in order to avoid the outcome of not winning a prize, given his high expectations regarding winning a prize. In general, the decrease in effort of low-ability contestants dominates the increase in effort of high-ability contestants. This may result in an overall decrease in the total expected effort. In this case, in order to compensate for the decrease in total expected effort, the contest designer motivates the low-ability contestants by introducing a second, or possibly a third or more prizes. This result is consistent with the experimental findings of Freeman and Gelber (2009). They experimentally study the effort provision in a real-effort tournament, where subjects are asked to solve mazes. In the experiment they implement different prize structures. They find that the number of solved mazes is higher when there are multiple differentiated prizes and that the number of solved mazes is lower when there is a single prize.[2]

My paper fits well into the recent and growing literature utilizing expectation-based loss aversion in different settings to give a rationale for a variety of empirical findings. Crawford and Meng (2011) analyze field data on cab drivers' working hours and propose a model of labor supply for cab drivers incorporating the K-R model. Their estimates suggest that their reference-dependent model of labor supply rationalizes the cab drivers' behavior observed in the field data. Herweg et al. (2010) study the principal agent model with moral hazard in the presence of expectation-based loss-averse agents. They show that the optimal contract is a binary payment scheme consistent with the observed prevalence of simple contracts. Lange and Ratan (2010) study first- and second-price sealed

[1] The intuition presented here is in line with the "loss contemplation" reasoning for overbidding in auctions presented in Delgado et al. (2008).

[2] M-S prove that multiple differentiated prizes might be optimal when the cost-of-effort is convex. Freeman and Gelber (2009) use solving mazes as a measure of effort provision in their experiment. In maze solving cost-of-effort is likely to be concave rather than convex since it becomes less costly as you solve more and more mazes due to learning.

bid auctions for a single item with expectation-based loss-averse bidders. Their model predicts overbidding in first-prize auctions, in line with evidence from recent laboratory experiments.[3]

In the remainder of this paper, I focus on the two-prize case for ease of exposition. I present the general results for equilibrium effort functions and the optimal prize allocation when there are $p > 2$ prizes in the Appendix. In Section 2.2, I present the model and in Section 2.3, I introduce further notation and discuss participation in the contest. In Section 2.4, I focus on linear cost-of-effort functions and derive the equilibrium effort of the contestants. Afterwards, I state the contest designer's problem and characterize the optimal prize allocation. I discuss the cases of convex and concave cost-of-effort functions in Section 2.5. I derive the optimal effort function of the contestants and provide a sufficient condition for the optimality of multiple prizes. Section 2.6 concludes. The proofs are relegated to the Appendix.

2.2. The Model

Consider a contest with p prizes $V_1 \geq V_2 \geq ... \geq V_p \geq 0$, where V_j denotes the value of the j-th prize. The values of the prizes are announced by the contest designer and are common knowledge. The prizes are normalized, so that $\sum_{i=1}^{p} V_i = 1$.

Furthermore, let there be k contestants, with $k \geq p$. Each contestant has an ability (cost) parameter c_i, which is private information. Ability parameters are drawn independently from a continuous distribution function F on the interval $[m, 1]$. The distribution function F is assumed to have a strictly positive and continuous density $F' > 0$. It is assumed that F is common knowledge.

All contestants simultaneously exert costly efforts. Denote contestant i's effort by x_i. Contestant i, exerting effort x_i, bears the cost-of-effort denoted by $c_i\gamma(x_i)$, where $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ is assumed to be a strictly increasing function with $\gamma(0) = 0$. Note that a high c_i means low ability (higher cost) for contestant i. In the remainder of the text, the contestants having higher c_is will be referred to as low-ability contestants and those with low c_is will be referred to as high-ability contestants. In order to avoid infinite efforts caused by zero costs, the highest possible ability m is assumed to be strictly positive.

The contestants are assumed to be expectation-based loss-averse in the sense of K-R. I will briefly introduce expectation-based loss aversion and explain how it translates into my model. According to K-R, the overall utility of an agent from consum-

[3] Kőszegi (2013) summarizes many other studies incorporating reference dependence preferences into theoretical models.

ing the n dimensional bundle $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$ when having the reference point $\mathbf{r} = (r_1, \ldots, r_n) \in \mathbb{R}^n$ is assumed to have two components: a consumption utility and a gain-loss utility. The consumption utility in dimension l is the standard outcome-based consumption utility and does not depend on the reference point. The gain-loss utility in dimension l captures how the agent feels about gaining and losing in this dimension. The gain-loss utility depends on how consumption in dimension l compares to agent's reference point. In particular, the overall utility of an agent from consuming $\mathbf{a} = (a_1, \ldots, a_n)$ when having the reference point $\mathbf{r} = (r_1, \ldots, r_n)$ is given by:

$$v(a|r) = \sum_{l=1}^{n} \upsilon_l(a_l) + \sum_{l=1}^{n} \mu(\upsilon_l(a_l) - \upsilon_l(r_l)) \tag{2.1}$$

Here, υ_l denotes the consumption utility in dimension l and μ denotes the gain-loss function. The gain-loss function is assumed to satisfy the assumptions Kahneman and Tversky (1979) put on their value function. In my framework, the consumption space of the contestant has two dimensions, that is $n = 2$: the prize dimension, i.e. $a_1 = V_j$ and the effort dimension, i.e. $a_2 = x_i$. I assume that the consumption utilities in both prize and effort dimensions are given by $\upsilon_j(.) = .$, for $j \in 1, 2$. Put verbally, the consumption utility of winning a prize V_j is identical to the value of that prize. Similarly, the consumption utility of exerting effort x_i is equal to the cost-of-effort $c_i\gamma(x_i)$. To discuss the gain-loss utility, it is first necessary to define the "gain-loss function" μ.

$$\mu(w) = \begin{cases} \eta w, & \text{if } w \geq 0 \\ \eta\lambda w, & \text{if } w < 0, \end{cases}$$

where $\lambda \geq 1$ is the weight attached to losses relative to gains and $\eta > 0$ is the weight attached to gain-loss utility relative to consumption utility. With this formulation, I assume a constant marginal utility from gains and a larger — in magnitude — marginal disutility from losses. In other words, losses loom larger than gains. However, $\mu(w)$ is not S-shaped in order to keep the analysis tractable.

According to K-R, the gain-loss utility is derived from the standard consumption utility and the reference point, as given in equation (2.1). The reference point is determined endogenously by the environment. I use personal equilibrium as the solution concept. Personal equilibrium states that the decision-maker must choose a state-contingent plan that is optimal given the preferences induced by the plan. That is, expectations should be consistent with optimal behavior given expectations. Given an outcome, the gain-loss utility is derived by comparing the given outcome to all possible outcomes that could have occurred and weighting each comparison with the *ex-ante* probability of the alternative outcome. The gain-loss utility for a given outcome is obtained by summing all these weighted comparisons. The utility from a given outcome is the sum of the standard consumption utility and the gain-loss utility. The expected utility of a contestant is the weighted average of all possible outcomes, given that the actual outcome itself is uncertain.

More precisely, suppose that there are two prizes to be awarded, $V_1 \geq V_2 \geq 0$, and $k > 2$ contestants. There are three possible outcomes for the contestant in this case: (i) winning first prize V_1, (ii) winning second prize V_2 and (iii) not winning any prize. Denote the probabilities with which these outcomes occur by p_1, p_2 and $(1 - p_1 - p_2)$, respectively. The outcome that contestant i wins first prize V_1 is evaluated as follows:

$$\underbrace{\underbrace{V_1}_{\text{consumption utility}} + \underbrace{\eta \left\{p_2 \left(V_1 - V_2\right) + \left(1 - p_1 - p_2\right) V_1\right\}}_{\text{gain-loss utility}}}_{\text{prize dimension}} + \underbrace{\underbrace{-c_i \gamma(x_i)}_{\text{consumption utility}} + \underbrace{0}_{\text{gain-loss utility}}}_{\text{effort dimension}} . \tag{2.2}$$

In this formulation, the first term is the consumption utility in the prize dimension, that is, the consumption utility from winning first prize, which is equal to the value V_1. The second term is the gain-loss utility in the prize dimension, which gives the contestant's feeling of gain or loss from winning first prize V_1. This term is obtained by comparing the given outcome - winning first prize - to all possible outcomes, namely winning second prize or not winning anything. Compared to the alternative outcome that the contestant ends up with second prize V_2, which happens with probability p_2, he experiences a gain of $V_1 - V_2$; meanwhile, compared to the alternative outcome where the contestant ends up not winning any prize, which happens with a probability $(1 - p_1 - p_2)$, he experiences a gain of V_1. The coefficient η is the weight of the gain-loss utility, which measures the weight attached to the gain-loss utility relative to the consumption utility. Note that in all these comparisons the contestant is in the gain domain, since winning first prize is the best outcome. The last term in 2.2 is the consumption utility in the effort dimension, namely the standard disutility of exerting effort x_i. The gain-loss utility in the effort dimension is simply zero, since the expected and the actual effort choices of the contestant coincide.

Similarly, the utility of contestant i from winning second prize V_2 is formulated as follows:

$$\underbrace{\underbrace{V_2}_{\text{consumption utility}} + \eta \underbrace{\{p_1 \lambda (V_2 - V_1) + (1 - p_1 - p_2) V_2\}}_{\text{gain-loss utility}}}_{\text{prize dimension}} + \underbrace{\underbrace{-c_i \gamma(x_i)}_{\text{consumption utility}} + \underbrace{0}_{\text{gain-loss utility}}}_{\text{effort dimension}}. \tag{2.3}$$

In the above evaluation, different from the first one, the loss aversion index λ comes into the picture. This is because the contestant is in the loss domain when he compares winning second prize V_2 to the alternative outcome of winning first prize V_1.

The utility of contestant i from not winning any prize is evaluated in the same way:

$$\underbrace{\underbrace{0}_{\text{consumption utility}} + \eta \underbrace{\{p_1 \lambda (-V_1) + p_2 \lambda (-V_2)\}}_{\text{gain-loss utility}}}_{\text{prize dimension}} + \underbrace{\underbrace{-c_i \gamma(x_i)}_{\text{consumption utility}} + \underbrace{0}_{\text{gain-loss utility}}}_{\text{effort dimension}}. \tag{2.4}$$

In comparisons of not winning any prize to the alternative outcomes of winning first and second prize, the contestant is in the loss domain. Note that not winning any prize is the least favorable outcome for the contestant, since each contestant bears the cost-of-effort regardless of winning a prize.

As the actual outcome is uncertain, the expected utility of contestant i with type c_i is given by the sum of (2.2), (2.3) and (2.4) weighted by their respective probabilities:

$$\begin{aligned} EU \quad = \quad & p_1\{V_1 + \eta(p_2(V_1 - V_2) + (1 - p_1 - p_2)V_1) - c_i\gamma(x_i)\} \\ & + p_2\{V_2 + \eta(p_1\lambda(V_2 - V_1) + (1 - p_1 - p_2)V_2) - c_i\gamma(x_i)\} \\ & + (1 - p_1 - p_2)\{\eta(p_1\lambda(-V_1) + p_2\lambda(-V_2))c_i\gamma(x_i)\}. \end{aligned} \tag{2.5}$$

Note that the probabilities p_1, p_2 and $(1 - p_1 - p_2)$ are affected by the effort that the contestant exerts: p_1 is the probability that the $(k-1)$ competitors of contestant i exerts less effort then contestant i and p_2 is the probability that $(k-2)$ competitors of contestant i exert less effort than him while one competitor exerts more effort. The probability of not winning any prize is given by $(1 - p_1 - p_2)$. Note that by changing his effort level, each contestant affects the probability of winning a prize as well the endogenous reference point. Letting $\lambda = 1$ and $\eta = 1$ equation (2.5) reduces to the expected utility under standard preferences as formulated in M-S.

The timing of the contest game is as follows. In the first stage, the contest designer chooses the number and the level of the prizes in order to maximize total expected effort. The designer's revenue is the sum of expected efforts. The prize sum is fixed and assumed to be the normalized $\sum_{i=1}^{k} V_i = 1$. In the second stage, given the prize structure, the contestants choose their effort levels in order to maximize their expected utility. The contestant with the highest effort wins first prize V_1, and the contestant with the second highest effort wins second prize V_2. In the case when all contestants exerts zero effort, no prize will be distributed. Each contestant bears the cost-of-effort regardless of winning any prize.

2.3. Participation in the Contest

Before discussing participation in the contest, it is convenient to introduce the following notation to ease the exposition. First, define $\Lambda = \eta(\lambda - 1)$, where η is the weight placed on the gain-loss utility relative to the consumption utility and λ is the degree of loss aversion. Λ is interpreted as an overall measure of an agent's degree of loss aversion (see also Herweg et al. (2010) and Eisenhuth and Ewers (2012)). Λ is strictly positive for a loss-averse agent while Λ equals zero with standard preferences. Rearranging the terms in equation (2.5) and substituting $\Lambda = \eta(\lambda - 1)$, the expected utility of contestant i can be rewritten as follows:

$$\begin{aligned} EU &= p_1 V_1 + p_2 V_2 - c_i \gamma(x_i) \\ &\quad - \Lambda \left\{ p_1 p_2 (V_1 - V_2) + (1 - p_1 - p_2)(p_1 V_1 + p_2 V_2) \right\}. \end{aligned} \tag{2.6}$$

Second, let $F_s(c)$, $s \in \{1, 2\}$, denote the probability that a contestant with type c has a higher type than $s - 1$ of his $k - 1$ competitors while he has a lower type than $k - s$ of his $k - 1$ competitors. To illustrate, $F_1(c)$ is the probability that all remaining $(k - 1)$ contestants have higher types, that is they are less able, and $F_2(c)$ is the probability that $(k - 2)$ of the remaining contestants have lower types while one of them has a higher type. In other words, F_1 and F_2 are the first- and second-order statistics. Recall that a low-ability contestant has a higher c_i leading to higher costs. Note that in equilibrium it is assumed that contestant i exerts higher effort than his competitors with higher types. Contestant i affects these probabilities of winning the first and the second prize by choosing his effort level x_i.

Now I will discuss the participation in the contest.[4] Note that when $\Lambda = 0$, the expected utility of the agent in equation (2.6) equals the expected consumption utility. In this case, the agent has standard preferences but no gain-loss sensation. M-S show that there is full participation in the contest under the assumption of standard preferences,

[4]Participation in the contest means undertaking positive efforts.

that is when $\Lambda = 0$. Whenever $\Lambda > 0$, the agent has the expected gain-loss utility, next to the expected consumption utility. Given the fact that first prize is always larger than or equal to second prize, the gain-loss utility – the second line of the equation (2.6)– is either zero or negative. Depending on the relative magnitudes of the gain-loss utility and the standard consumption utility, the agent may end up with negative expected utility. Put differently, the agent has a non-negative expected utility only if the expected gain-loss utility does not dominate the expected consumption utility. If the agent is sufficiently loss-averse, that is when Λ is sufficiently large, he may end up with negative expected utility whenever he exerts positive effort. In order to avoid this situation, he exerts zero effort and stays out of the contest. Intuitively, whenever loss aversion is too pronounced, the primary concern of a contestant with a low probability of winning becomes reducing the likelihood of possible losses. In this case, he gives up the slim hope of winning a prize and avoids losses by reducing his effort level to zero.

Rearranging the terms in the expected utility given by equation (2.6), I obtain a condition that guarantees a contestant's participation in the contest. A contestant with ability parameter c derives a non-negative expected utility from participating in the contest if and only if:

$$\frac{F_1(c)^2V_1 + 2F_1(c)F_2(c)V_2 + F_2(c)^2V_2}{F_1(c)V_1 + F_2(c)V_2} > 1 - \frac{1}{\Lambda}. \tag{2.7}$$

Note that whenever $\Lambda \leq 1$, the condition in (2.7) is satisfied for any parameter $c \in [m, 1]$, implying that each contestant has a nonnegative expected utility. However, whenever $\Lambda \geq 1$, condition (2.7) may be violated for some contestants with sufficiently small probabilities of winning a prize. Therefore, we obtain:

Lemma 2.1. *There is full participation in the contest when $\Lambda \leq 1$. When $\Lambda > 1$, there is a critical type $\tilde{c}$ satisfying (2.7) with equality such that contestants with the ability $c > \tilde{c}$ drop-out by exerting zero effort.*

Lemma 2.1 guarantees full participation in the contest whenever $\Lambda \leq 1$ (see also Herweg et al. (2010), Eisenhuth and Ewers (2012)). Put differently, when players are sufficiently loss averse, i.e. $\Lambda > 1$, there is a group of players who exert zero effort and do not participate in the contest. This result is consistent with the recent experimental evidence (see Müller and Schotter (2010), Barut and Noussair (2002), Noussair and Silver (2006), Klose and Sheremeta (2012), Ernst and Thöni (2009)).

2.4. Linear Cost Functions

In this section, I will solve the contestants' and the designer's problems, respectively, for the linear cost-of-effort function. I will first derive the optimal behavior of the contestants for a given prize structure. Next, given the optimal behavior of the contestants for any prize structure, I will characterize the optimal prize allocation.

2.4.1. Contestants' Problem

Assume that the contestants have linear cost-of-effort functions, that is $\gamma(x) = x$. The following proposition displays the equilibrium effort function of a contestant when there are two prizes to be awarded and there are $k > 2$ loss-averse contestants.

Proposition 2.1. *Assume that there are two prizes $V_1 \geq V_2 \geq 0$ to be awarded and $k > 2$ contestants. If $\Lambda > 1$, then there exists a critical type $\tilde{c}$ satisfying (2.7) with equality, such that in equilibrium contestants with $c \geq \tilde{c}$ exert zero effort and contestants with $c < \tilde{c}$ exert effort according to:*

$$b(c) = A(c)V_1 + B(c)V_2 \tag{2.8}$$

where the coefficients of the first and second prize are given by:

$$A(c) = (1 - \Lambda) \int_c^{\tilde{c}} -\frac{1}{a} F_1'(a) da + \Lambda \int_c^{\tilde{c}} -\frac{1}{a} (F_1^2(a))' da \tag{2.9}$$

and

$$B(c) = (1 - \Lambda) \int_c^{\tilde{c}} -\frac{1}{a} F_2'(a) da + \Lambda \int_c^{\tilde{c}} -\frac{1}{a} \left((F_2^2(a))' + (2F_1(a)F_2(a))' \right) da. \tag{2.10}$$

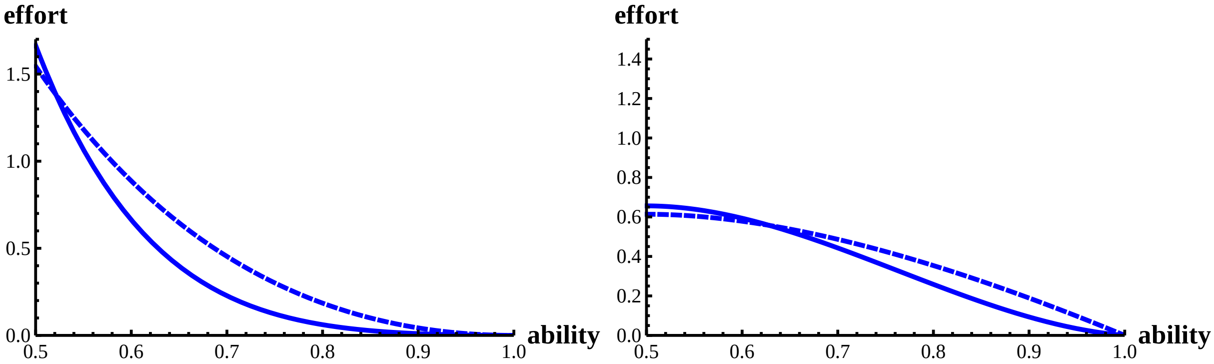
If $\Lambda \leq 1$, then each contestant exerts effort according to equation (2.8), where $A(c)$ and $B(c)$ are as in equations (2.9) and (2.10) with $\tilde{c} = 1$.

Proof. See Appendix 2.A. *Q.E.D.*

The equilibrium effort function for the general case with p prizes and $k \geq p$ contestants is derived in Appendix C.

By Lemma 2.1, full participation in the contest is guaranteed when $\Lambda \leq 1$. In equilibrium, each contestant exerts an effort equal to a weighted sum of first and second prize. The weights of the prizes differ for each contestant depending on his chances of winning first and second prize. When $\Lambda > 1$, there is a subset of contestants who exert 0 effort in equilibrium. Note that by letting $\Lambda = 0$, the above equilibrium effort functions reduce to those with standard preferences formulated in M-S.

Figure 2.1: Equilibrium Effort Functions



(a) Single prize, $V_1 = 1$ and $V_2 = 0$. (b) Two equal prizes, $V_1 = V_2 = 0.5$.

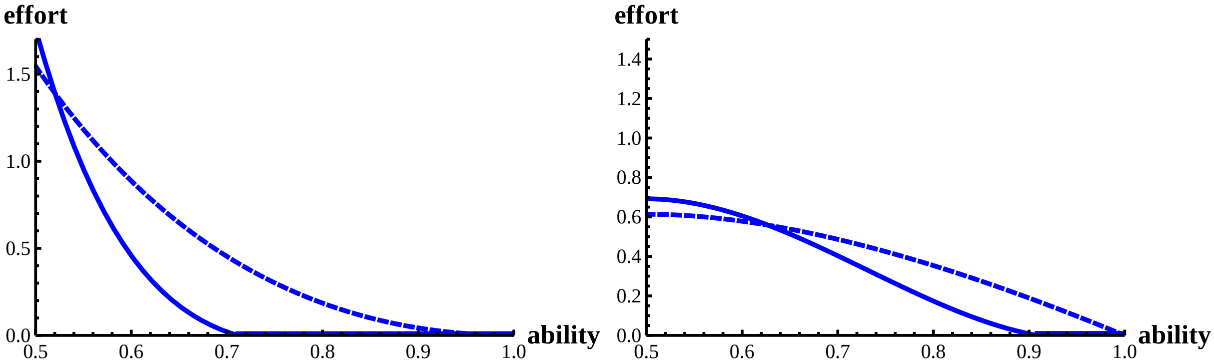
Notes: The left panel depicts the equilibrium effort functions when the designer awards a single prize. The right panel depicts the equilibrium effort functions when the designer awards two equal prizes. The degree of loss aversion of the contestants is $\Lambda = 0.8$.

The following example illustrates the equilibrium effort function of contestants under a uniform distribution of abilities.

Example 2.1. Assume that there are $k = 3$ contestants whose abilities are drawn from the uniform distribution $F(c) = 2c - 1$ on the interval $[1/2, 1]$. First, let $\Lambda = 0.8$, guaranteeing that each contestant participates in the contest (see Lemma 2.1). Figure 2.1 depicts the equilibrium effort function in the presence of standard preferences (dashed line) and expectation-based reference-dependent preferences (solid line).

Recall that the expected utility of a contestant (see equation 2.6) has two parts: the expected consumption utility and the expected gain-loss utility. The expected consumption utility is equal to the expected utility of a contestant with standard preferences. Therefore, the difference in the equilibrium behavior between a contestant with standard preferences and a contestant with reference-dependent preferences stems from the expected gain-loss utility. Recall that the gain-loss utility is derived by comparing the actual outcome with the contestant's expectations. This is the point where expectations come into the picture. A contestant with high ability has an *ex-ante* high probability of winning a prize and has therefore high expectations. In order to avoid the loss of not winning a prize, he increases his probability of winning by increasing his effort level further. On the other hand, a contestant with low ability has an *ex-ante* low probability of winning a prize, and has therefore low expectations. In order to reduce the scope of possible losses, he reduces his expectations further by lowering his effort level. Therefore, expectation-based loss aversion incentivizes high-ability contestants to increase their effort level while it induces low-ability contestants to lower their effort levels. Therefore, high-ability contestants exert more effort while low-ability ones exert less effort in com-

Figure 2.2: Equilibrium Effort Functions



(a) Single prize, $V_1 = 1$ and $V_2 = 0$.

(b) Two equal prizes, $V_1 = 1 = V_2 = 0.5$.

Notes: The left panel depicts the equilibrium effort functions when the designer awards a single prize. The right panel depicts the equilibrium effort functions when the designer awards two equal prizes. The degree of the loss aversion of the contestants is $\Lambda = 1.5$.

parison to the predictions of standard preferences.

It is important to note that there are two different ways through which contestants try to avoid losses. One is by increasing the effort level in order to increases the chances of winning a prize, and the other one is by decreasing the effort level in order to lower expectations. High-ability contestants use the first the way since their ex-ante chances of winning a prize is already high. Low-ability contestants use the second way and decrease their effort level to decrease their expectations. This is because if a low-ability contestant tries to put more effort, he faces higher cost of effort in comparison to a high-ability contestant.

Now let $\Lambda = 1.5$, in which case there is a critical type $\tilde{c}$ satisfying condition (2.7) with equality such that any type $c \geq \tilde{c}$ exerts zero effort by Lemma 2.1. Figure 2.2 depicts the equilibrium effort functions when $\Lambda = 1.5$.

When the overall degree of loss aversion Λ exceeds 1, we still see the aggressive effort provision of high-ability contestants and the under-exertion of effort of low-ability contestants. In addition to these findings, the dropping-out behavior of low-ability contestants occurs. Intuitively, when a low-ability contestant is sufficiently loss-averse, the gain-loss utility dominates the standard consumption utility. In this case, the contestant focuses on reducing the net loss arising from the gain-loss utility and exerts zero effort. These results are consistent with the experimental evidence presented in Müller and Schotter (2010).

2.4.2. Designer's Problem

Given the optimal behavior of contestants for any prize allocation, the contest designer chooses the number and the level of the prizes. The goal of the contest designer is to maximize his expected revenue, namely the total expected effort exerted by contestants. Let $V_2 = \alpha$ and $V_1 = 1 - \alpha$, where $0 \leq \alpha \leq 1/2$.

Recall that whenever $\Lambda > 1$, there is a positive mass of types $c \geq \tilde{c}$ exerting zero effort by Lemma 2.1. Contestants with $c < \tilde{c}$ exert effort according to equation (2.8). When $\Lambda \leq 1$, there is full-participation, so that $\tilde{c} = 1$. The average effort of each contestant is given by:

$$\int_m^{\tilde{c}} b(c)F'(c)dc = \int_m^{\tilde{c}} (1-\alpha)A(c) + \alpha B(c)F'(c)dc, \tag{2.11}$$

where $A(c)$ and $B(c)$ are given by equations (2.9) and (2.10). As there are k contestants, the designer's problem is given by:

$$\max_{0\leq\alpha\leq 1/2} k \int_m^{\tilde{c}} (A(c) + \alpha(B(c) - A(c)))F'(c)dc. \tag{2.12}$$

Since the maximization is over α, the designer's problem can be written as follows:

$$\max_{0\leq\alpha\leq 1/2} \alpha \int_m^{\tilde{c}} (B(c) - A(c))F'(c)dc. \tag{2.13}$$

The solution to the designer's problem depends on the sign of the integral in equation (2.13): it is optimal to award a single prize if the integral is negative, and to award two equal prizes otherwise. Note that awarding two unequal prizes is never optimal due to the linearity of the program. The sign of the integral depends on the specific properties of the distribution function F of abilities, the number of contestants k and the degree of loss aversion Λ.

Proposition 2.2. *Assume that there are at most two prizes to be awarded with $V_1 \geq V_2 \geq 0$ and $k > 2$ contestants with linear cost-of-effort functions. Then it is optimal to allocate the whole prize sum to a single prize if, and only if:*

$$\int_m^{\tilde{c}} (B(c) - A(c))\, F'(c)dc < 0 \tag{2.14}$$

and to award two equal prizes otherwise.

Proof. See Appendix 2.B. *Q.E.D.*

The solution to the designer's problem for the general case with p prizes, $k \geq p$ contestants and any Λ is derived in Appendix 2.D. The following example illustrates the optimal prize allocation under a uniform distribution of abilities.

Example 2.2. Assume that there are 3 contestants, whose abilities are drawn from a uniform distribution $F(c) = 2c - 1$ on the interval $[0.5, 1]$. Figure 2.3 depicts the equilibrium effort functions when the designer announces a single grand prize, $b^{(1,0)}$, and two equal prizes, $b^{(0.5,0.5)}$ separately. The indices $(1, 0)$ and $(0.5, 0.5)$ refer to the prize allocations $V_1 = 1$, $V_2 = 0$ and $V_1 = 0.5$, $V_2 = 0.5$, respectively. The dashed and the bold lines are the equilibrium effort functions under the assumption of standard preferences and expectation-based reference-dependent preferences, respectively.

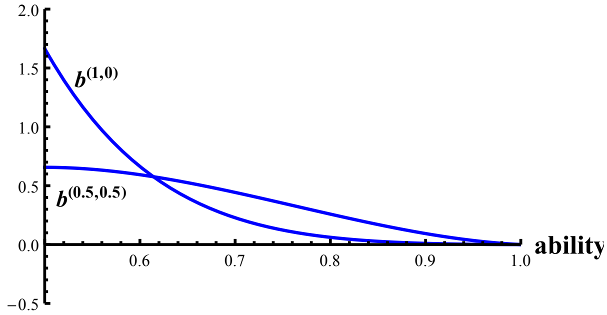
In general — for both preference types — a second prize motivates low-ability contestants to increase their effort level. Intuitively, low-ability contestants would give up the competition if there is only a single prize and exert more effort when the contest designer announces a second prize. On the other hand, a second prize will give high-ability contestants an incentive to lower their effort levels. This is because high-ability contestants are mainly competing for first prize and introducing a second prize will lower the value of the first (since the prize sum is constant). Figure 2.3 illustrates the effort decrease of high-ability contestants and the effort increase of low-ability ones in the presence of a second prize.

The contest designer decides on whether to introduce a second prize by comparing the differences in effort provision of high and low-ability contestants. If the increase in total expected effort by low-ability contestants — in the presence of a second prize — dominates the decrease in total expected effort by high-ability contestants, then the contest designer is better off by introducing a second prize.

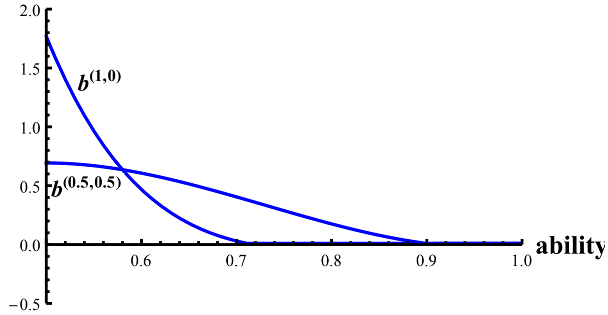
M-S show that when contestants have standard preferences, the effort increase of low-ability contestants does not compensate for the effort decrease of high-ability contestants relative to the single prize case, so that a single first prize is optimal. A reasonable conjecture is that the result of this comparison will depend on the number of contestants and the specific properties of the ability distribution. Surprisingly M-S show that this conjecture is wrong in their setup, i.e. their result does not depend on these variables. When contestants have expectation-based reference-dependent preferences, however, the comparisons of effort provision across types depend on the variables.

For the specific values taken in this example, and contrary to the case of standard preferences, it is optimal to award two equal prizes. The reason is that loss aversion leads low-ability contestants to provide little or no effort and high-ability contestants to exert effort aggressively in comparison to the predictions of standard preferences. In this case, the effort increase of low-ability contestants does compensate for the effort decrease of high-ability contestants relative to the single prize case. As such, the contest designer is better off when he allocates the total prize sum as two equal prizes. The optimality of two equal prizes - rather than two unequal prizes - is due to the linearity

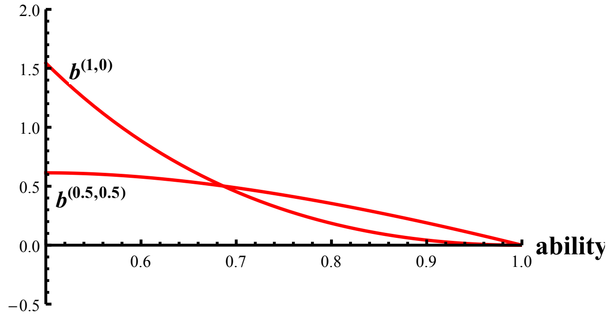
Figure 2.3: The Beneficial Effect of Second Prize



(a) $\Lambda = 0.8$.



(b) $\Lambda = 1.5$.



(c) $\Lambda = 0$.

Notes: The figure depicts the optimal effort functions in the presence of a single and two prizes. The graphs on the upper panel are the equilibrium effort functions in the presence of reference dependent preferences preferences and the one on the lower panel is in the presence of standard preferences preferences.

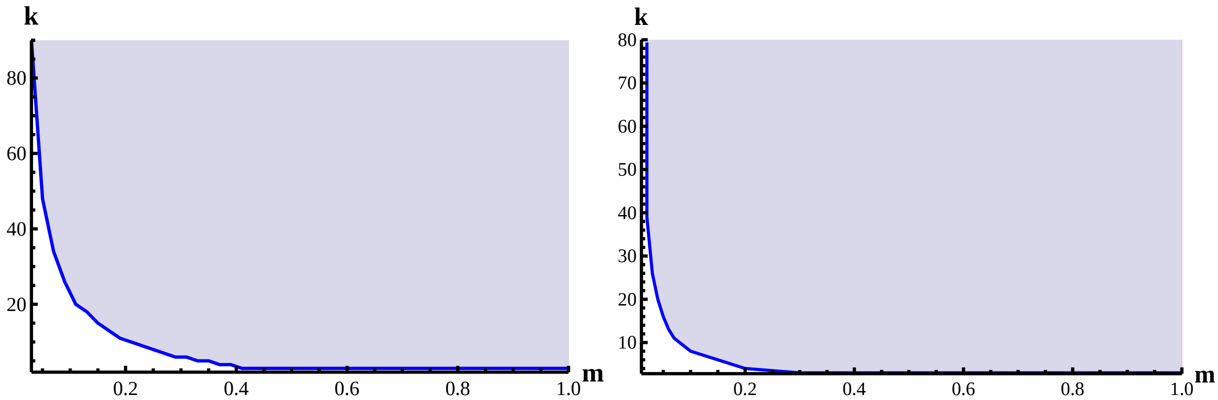
of the program (see the proof of Proposition 2.2). When $\Lambda = 1.5$, the effort decrease of low-ability contestants becomes more prominent due to drop-outs, depicted in the right panel of Figure 2.3.

Figure (2.4) depicts the optimal prize structure for the combination of different values for k and m under a uniform distribution of abilities. For the values in the shaded area it is optimal to award two equal prizes, while a single prize is optimal in the unshaded area. As the overall degree of loss aversion increases, the area over which two equal prizes are optimal expands.

As the number of contestants k increases, keeping everything else constant, the beneficial effect of a second prize on the total expected effort increases. Intuitively, a contestant has a lower probability of winning when there are more competitors. All but the high-ability contestants will have lower expectations regarding winning a prize if there are more competitors. The contest designer motivates these contestants by introducing a second prize, allowing him to obtain a higher total expected effort.

As the minimum effort cost m increases it becomes optimal to award two prizes. One

Figure 2.4: Optimal Prize Allocation



(a) $\Lambda = 0.8$.

(b) $\Lambda = 1.5$.

Notes: The figure illustrates the optimal allocation of prizes depending on the number of contestants k and the lowest type m. For the values of k and m in the unshaded area, it is optimal to award a single prize, while for the values in the shaded area it is optimal to allocate the total prize sum as two equal prizes.

way to explain this result is as follows. In the case where m is large contestants have high cost of efforts in comparison to the case of a small m. So, the number of contestants who overexert effort will be much less in the case of a large m in comparison to the case of a small m. When m is small, the overexertion of effort by high-ability contestants compensate for the under exertion of effort and dropping out behavior. In this case the contest is better off by awarding a single grand prize and motivating the high-ability contestants. While, when m is large, the reasoning goes in the opposite direction since there is relatively less number of contestants who overexert effort. In this case the contest designer is better off by awarding two prizes and motivating low-ability contestants.

2.5. Concave and Convex Cost Functions

In this section, I will solve the contestants' and the designer's problem, respectively, for convex or concave cost-of-effort functions, similar to the previous section. I will first derive the optimal behavior of the contestants for a given prize structure. Next, given the optimal behavior of the contestants for any prize structure, I will characterize the optimal prize allocation.

2.5.1. Contestants' Problem

Assume that the contestants have either concave or convex cost-of-effort functions with $\gamma(0) = 0$ and γ being an increasing function. The following proposition displays the equilibrium effort function of a contestant when there are two prizes to be awarded and there are $k > 2$ contestants.

Proposition 2.3. *Assume that there are two prizes $V_1 \geq V_2 \geq 0$ to be awarded and $k > 2$ contestants. If $\Lambda > 1$, then there exists a critical type $\tilde{c}$ satisfying (2.7) equality such that – in equilibrium – contestants with $c \geq \tilde{c}$ exert zero effort and contestants with $c < \tilde{c}$ exert effort according to:*

$$b(c) = \gamma^{-1}\left(A(c)V_1 + B(c)V_2\right), \tag{2.15}$$

where the coefficients of first and second prize are given by equations (2.9) and (2.10), respectively. If $\Lambda \leq 1$, then the optimal effort for all types is positive and given by equation (2.15), where $A(c)$ and $B(c)$ are defined by equations (2.9) and (2.10) with $\tilde{c} = 1$.

Proof. See Appendix 2.A. *Q.E.D.*

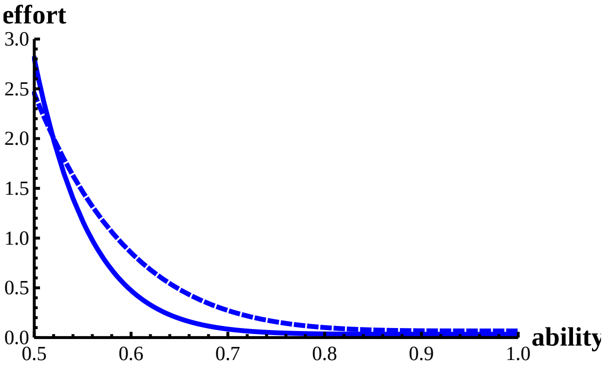
The equilibrium effort function for the general case with p prizes and $k \geq p$ contestants is derived in Appendix 2.C. The equilibrium effort of each contestant is given by a simple transformation of the equilibrium effort obtained in the linear cost case. Note that when $\Lambda = 0$, the equilibrium above reduces to that with standard preferences formulated in M-S.

The following example illustrates the equilibrium effort function of contestants with convex and concave cost-of-effort functions, respectively.

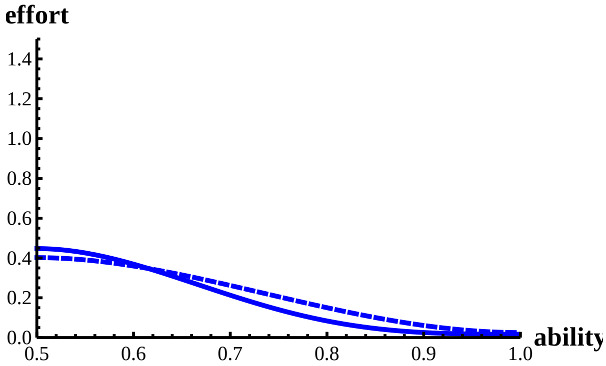
Example 2.3. Assume that there are $k = 3$ contestants, whose abilities are drawn independently from the uniform distribution $F(c) = 2c - 1$ on the interval $[1/2, 1]$, as in example 2.1. Assume that the concave cost-of-effort function is $\gamma(x) = \sqrt{x}$ and the convex cost-of-effort function is $\gamma(x) = x^2$. Figure 2.5 and 2.6 depict the equilibrium effort functions when contestants have concave and convex cost-of-effort functions, respectively. The upper and lower panels of these figures illustrate the effort provision in equilibrium respectively in the cases where there is full participation in the contest (when $\Lambda = 0.8$) and there is dropping out (when $\Lambda = 1.5$).

The equilibrium effort functions in the case of convex or concave cost-of-effort functions is obtained by a simple transformation of the equilibrium effort curve found in the linear cost-of-effort case. Therefore, the intuition provided in Example 2.1 applies to the cases of concave or convex cost-of-effort functions in the same way. In particular, high-ability contestants aggressively exert effort while low-ability contestants exert little or no effort, relative to the predictions with standard preferences. This is because a contestant with high ability, holding high expectations, exerts effort aggressively in order to

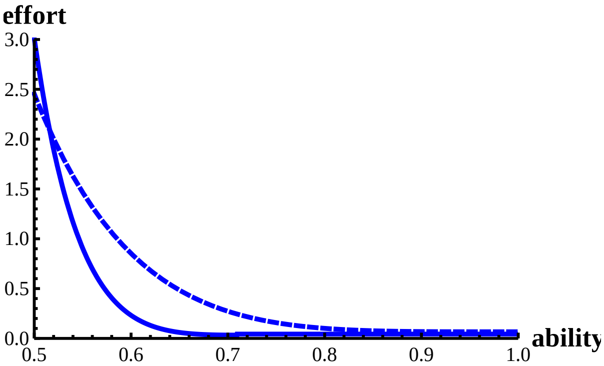
Figure 2.5: Equilibrium Effort Functions for Concave Costs $\gamma(x) = \sqrt{x}$



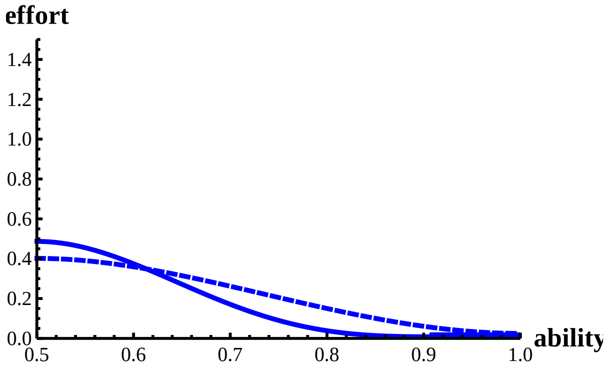
(a) Single prize, $V_1 = 1$ and $V_2 = 0$.



(b) Two equal prizes, $V_1 = 1 = V_2 = 0.5$.



(c) Single prize, $V_1 = 1$ and $V_2 = 0$.



(d) Two equal prizes, $V_1 = 1 = V_2 = 0.5$.

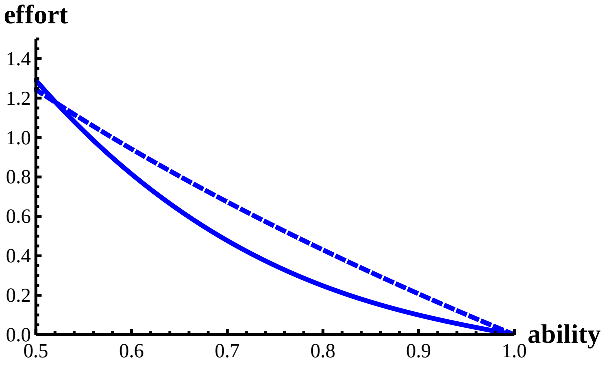
Notes: The left panels depict the equilibrium effort curves when there is a single prize, while the right panels depict the equilibrium effort curves when there are two equal prizes. The upper and the lower panels illustrate the equilibrium effort curves, respectively, for $\Lambda = 0.8$ and $\Lambda = 1.5$.

avoid the loss of not winning a prize. On the other hand, a contestant with low ability, holding low expectations, exerts little effort to reduce his expectations further in order to minimize the loss sensation stemming from their gain-loss utility. Whenever contestants are sufficiently loss-averse, i.e. $\Lambda > 1$, low-ability contestants exert zero effort, dropping out of the contest. The reason is that the gain-loss utility might dominate the standard consumption utility for a low-ability contestant. In this case, the contestant's primary concern becomes avoiding possible losses, incentivizing him to drop his effort level to zero.

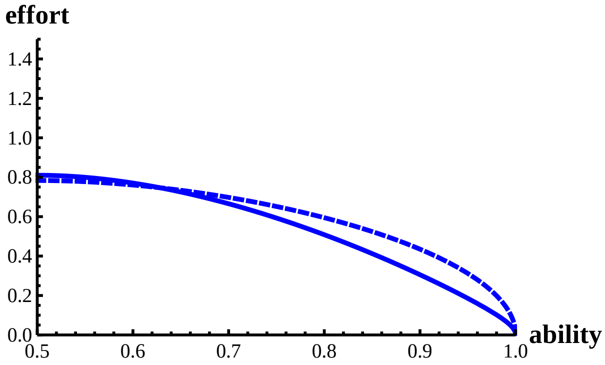
2.5.2. Designer's Problem

Let $V_2 = \alpha$ and $V_1 = 1 - \alpha$, where $0 \leq \alpha \leq 1/2$. Analogous to the case of linear cost-of-effort functions, the average effort of each contestant with a convex or concave cost-of-effort function is given by:

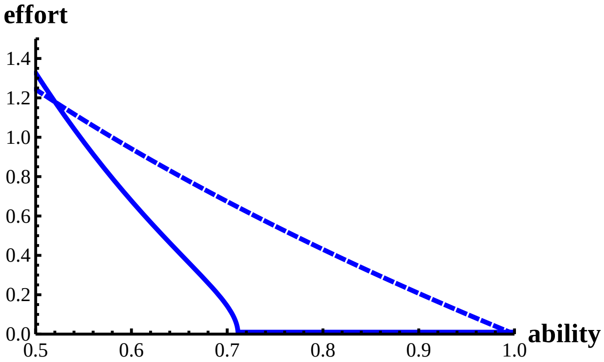
Figure 2.6: Equilibrium Effort Functions for Convex Costs $\gamma(x) = x^2$



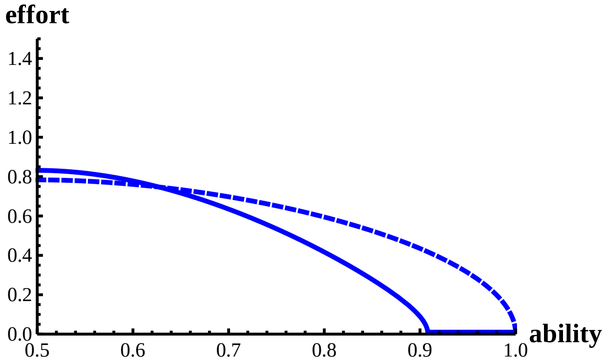
(a) Single prize, $V_1 = 1$ and $V_2 = 0$.



(b) Two equal prizes, $V_1 = 1 = V_2 = 0.5$.



(c) Single prize, $V_1 = 1$ and $V_2 = 0$.



(d) Two equal prizes, $V_1 = 1 = V_2 = 0.5$.

Notes: The left panels depict the equilibrium effort curves when the designer awards a single prize. The right panels depict the equilibrium effort curves when the designer awards two equal prizes. For both structures, the degree of loss aversion of the contestants is $\Lambda = 1.5$.

$$\int_m^{\tilde{c}} \gamma^{-1} \left(A(c) + \alpha(B(c) - A(c))\right) F'(c)dc \tag{2.16}$$

where $A(c)$ and $B(c)$ are given by equations (2.9) and (2.10). Note that whenever $\Lambda \leq 1$, full participation in the contest is guaranteed (see Lemma 2.1) so that $\tilde{c} = 1$. Since there are k contestants, the total expected effort — the revenue of the designer — is given by:

$$R(\alpha) = k \int_m^{\tilde{c}} \gamma^{-1}(A(c) + \alpha(B(c) - A(c)))F'(c)dc. \tag{2.17}$$

Since the goal of the designer is to maximize the total expected effort, the designer's problem becomes:

$$\max_{0 \leq \alpha \leq 1/2} k \int_m^{\tilde{c}} \gamma^{-1}(A(c) + \alpha(B(c) - A(c)))F'(c)dc. \tag{2.18}$$

The solution to the designer's problem depends on the shape of the revenue function $R(\alpha)$. More specifically, awarding a single prize is optimal if $R(\alpha)$ is strictly decreasing, that is if the revenue function has its maximum at $\alpha = 0$. Otherwise, the revenue function $R(\alpha)$ might have its maximum at $\alpha \neq 0$, leading to the optimality of the two prizes. The shape of the revenue function $R(\alpha)$ depends on the degree of loss aversion Λ as well as the number of contestants and the specific properties of the distribution function F. If the shape of the revenue function $R(\alpha)$ is concave, the maximization problem of the designer might have an interior solution with $\alpha^* \in (0, 1/2)$. In this case, two unequal prizes become optimal, in contrast to the case of linear cost-of-efforts. In the following proposition, I provide a sufficient condition for the optimality of two prizes.

Proposition 2.4. *Assume that there are at most two prizes to be awarded with $V_1 \geq V_2 \geq 0$ and $k > 2$ contestants with convex or concave cost-of-effort functions. A sufficient condition for the optimality of two prizes is given by:*

$$\int_m^{\tilde{c}} (B(c) - A(c))g'(A(c))F'(c)dc > 0. \tag{2.19}$$

If condition (2.19) is satisfied, then it is optimal to award two prizes $V_1 = 1 - \alpha^*$ and $V_2 = \alpha^*$ with $R'(\alpha^*) = 0$, otherwise it is optimal to award a single prize.

Proof. See Appendix 2.B. *Q.E.D.*

Letting $\Lambda = 0$, the condition (2.19) reduces to that provided in M-S. The integral in condition (2.19) is an increasing function of the number of competitors. Hence if the number of competitors is high enough, then it is optimal to award two prizes. The ratio of the prizes depends on the distribution of types as well as their degree of loss aversion.

If the cost-of-effort is concave and there is full participation in the contest - that is if $\Lambda \leq 1$ - then the shape of the revenue function $R(\alpha)$ is convex. In this case, the maximization problem in equation (2.18) has corner solutions. In other words, it is optimal to award either a single prize or two equal prizes, obtaining the following corollary:

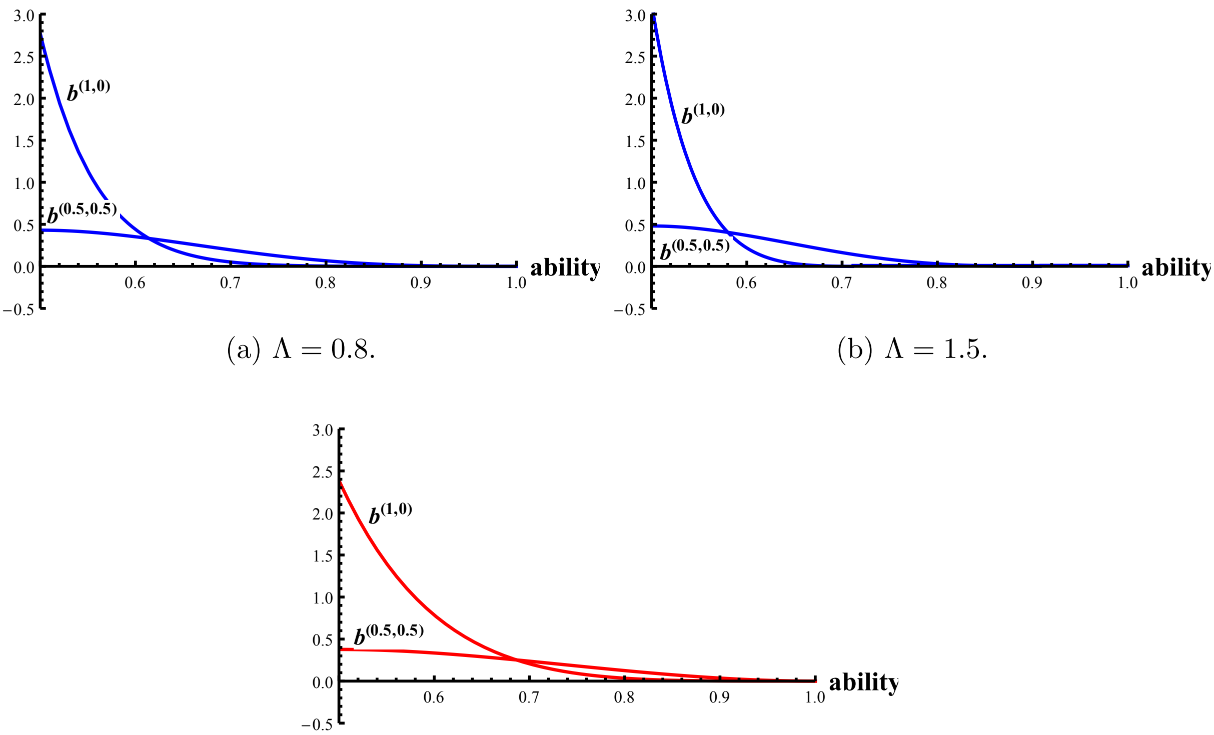
Corollary 2.1. *Assume that there are at most two prizes to be awarded with $V_1 \geq V_2 \geq 0$ and $k > 2$ contestants with concave cost-of-effort functions. If $\Lambda \leq 1$, then it is optimal to award either a single prize or two equal prizes.*

Proof. See Appendix 2.B. *Q.E.D.*

The following example illustrates the optimal prize allocation for concave cost-of-efforts under a uniform distribution of abilities.

Example 2.4. Assume that there are k contestants, whose abilities are drawn from a uniform distribution $F(c) = 2c - 1$ on the interval $[1/2, 1]$. Assume, moreover, that the

Figure 2.7: The Beneficial Effect of Second Prize



(a) $\Lambda = 0.8$.

(b) $\Lambda = 1.5$.

(c) $\Lambda = 0$.

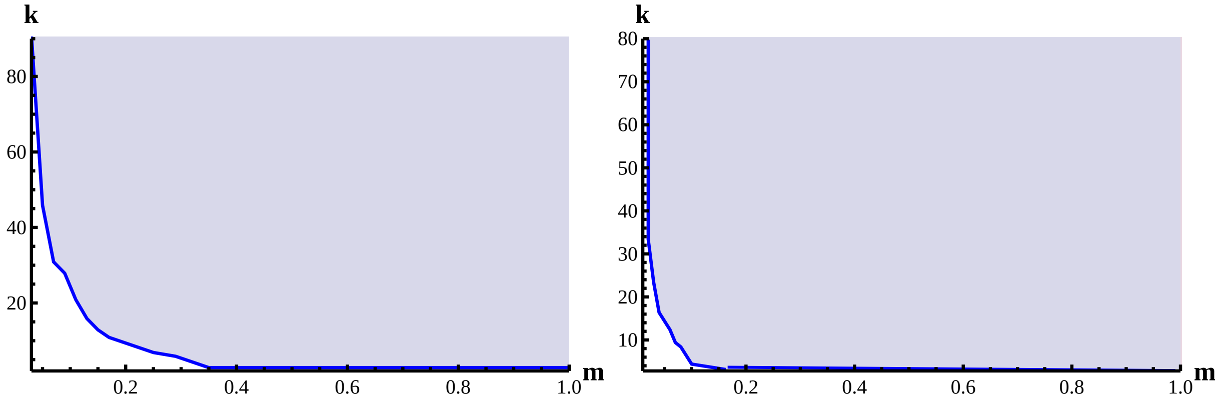
Notes: The figure depicts the optimal effort functions in the presence of a single and two prizes. The graphs on the upper panel are the equilibrium effort functions in the presence of reference dependent preferences preferences and the one on the lower panel is in the presence of standard preferences preferences.

cost-of-effort function is $\gamma(x) = \sqrt{x}$. Figure 2.7 depicts the equilibrium effort functions in the case of a single prize, $b^{(1,0)}$, and two equal prizes, $b^{(0.5,0.5)}$. The dashed and the solid lines are the equilibrium effort curves under the assumption of standard preferences and expectation-based reference-dependent preferences, respectively.

Since the equilibrium effort curve in the case of a concave cost-of-effort function is a transformation of that obtained in the case of a linear cost-of-effort function, the intuition presented in Example 2.2 applies to this example as well. Particularly, introducing a second prize motivates low-ability contestants to increase their effort levels while leading high-ability contestants to lower their effort levels. Figure 2.7 illustrates the decrease in effort of high-ability types and the increase in effort of low-ability types, in the presence of a second prize. If the former effect compensates for the latter one, it is optimal to award a second prize.

M-S show that – when contestants have standard preferences – it is optimal to award a single prize in the case of concave cost-of-effort functions. As in the case of linear cost-of-effort, they show that this prediction is independent of the number of contestants and

Figure 2.8: Optimal Prize Allocation



(a) $\Lambda = 0.8$. (b) $\Lambda = 1.5$.

Notes: The figure illustrates the optimal allocation of prizes depending on the number of contestants k and the lowest type m. For the values of k and m in the unshaded area, it is optimal to award a single prize. When $\Lambda = 0.8$ for the values of k and m in the unshaded area, it is optimal to award two equal prizes, and when $\Lambda = 1.5$ it might be optimal to allocate the prize sum as two unequal prizes.

the ability distribution. When contestants have expectation-based reference-dependent preferences, however, awarding a second prize can be optimal depending on the number of players and the ability distribution. Figure (2.8) depicts the optimal prize structure for different values of k and m under a uniform distribution of abilities.

Figure 2.8a illustrates the case with full participation in the contest. In this case, it is optimal to award two equal prizes for the values of k and m in the shaded area, and to award a single prize in the remaining area. In comparison to the linear cost-of-effort functions, the optimality of a single prize becomes less likely. This is because, with the concave cost-of-effort functions, the ability range over which contestants exert little effort is larger relative to linear cost-of-effort functions. Figure 2.8b illustrates the case in which low-ability contestants drop out. In this case, the beneficial effect of a second prize becomes more prominent for the contest designer, so that the area over which it is optimal to offer a single prize shrinks. In contrast to the case of linear cost-of-effort, it can be optimal to award two unequal prizes when there is dropping-out behavior. As the values of k and m increase, awarding two prizes becomes optimal, as discussed in Example 2.2.

2.6. Conclusion

In this paper, I studied a multiple prize contest under incomplete information, generalizing the contest model of Moldovanu and Sela (2001) by allowing for expectation-based loss aversion according to Kőszegi and Rabin (2006). The model presented in this paper is able to align the common experimental finding that high-ability contestants exert effort aggressively while low-ability contestants exert very little effort or drop out of the contest, in comparison to the predictions with standard preferences. An expectation-based loss-averse contestant has an expected gain-loss utility next to his expected consumption utility. The expected gain-loss utility measures the net loss sensation derived by comparing the actual outcome to all other alternative outcomes that might have occurred. High-ability and low-ability contestants have different incentives in order to avoid the feeling of loss stemming from the gain-loss utility. Intuitively, a high-ability contestant, who has high expectations for winning a prize, increases his effort level in order to avoid the loss of not winning a prize. A low-ability contestant, who has low expectations for winning a prize, decreases his expectations further by exerting very little effort to avoid the situation of losing a prize. When loss aversion is sufficiently pronounced, gain-loss utility dominates the standard consumption utility. In this case, and in order to avoid the net loss, a low-ability contestant exerts zero effort and drops out of the contest. The second main result is that in the presence of expectation-based loss aversion, awarding multiple prizes can be optimal where standard preferences predict the optimality of a single prize. The beneficial effect of a second prize becomes more prominent when contestants are expectation-based loss-averse. The reason is that low-ability contestants provide little effort due to their low expectations regarding winning a prize. The contest designer can increase his revenue — total expected effort — by motivating low-ability contestants with a second or possibly a third or more prizes. The optimality of multiple prizes is consistent with the prevalence of multiple prize contests in the real world.

Appendices

2.A. Derivation of Equilibria

Proof of Proposition 2.1. Assume that all contestants except i exert effort according to the function b. Moreover assume that b is strictly monotonic and differentiable. I will derive the optimal effort function first for the case when there is full participation in the contests (when $\Lambda \leq 1$) and then for the case when some contestants drop out the contest (when $\Lambda > 1$).

Suppose that each contestant participates in the contest, that is $\Lambda \leq 1$. The maximization problem of the contestant i is:

$$\begin{aligned} \max x \quad & \{p_1\{V_1 + \eta(p_2(V_1 - V_2) + (1 - p_1 - p_2)V_1) - cx\} \\ & + \ p_2\{V_2 + \eta(p_1\lambda(V_2 - V_1) + (1 - p_1 - p_2)V_2) - cx)\} \\ & + \ (1 - p_1 - p_2)\{\eta(p_1\lambda(-V_1) + p_2\lambda(-V_2)) - cx\}\}. \end{aligned} \tag{2.20}$$

where the probabilities of winning the first and the second prize ,p_1 and p_2, are defined as

$$\begin{aligned} p_1 =& (1 - F(b^{-1}(x)))^{k-1} \\ p_2 =& (k - 1)(1 - F(b^{-1}(x)))^{k-2}F(b^{-1}(x)). \end{aligned} \tag{2.21}$$

p_1 is the probability that all remaining $(k-1)$ contestants have higher types, that is they are less able, and p_2 is the probability that $(k - 2)$ of the remaining contestants have lower types while one of them has a higher type. Note that a contestant affects these probabilities of winning the first and the second prize by choosing his effort level x.

Denote the inverse effort function b^{-1} by y. Substituting b^{-1} and $\Lambda = \eta(\lambda - 1)$ and rearranging the terms, the maximization problem becomes:

$$\begin{aligned} max_x \quad & \big\{(1 - \Lambda)(1 - F(y))^{k-1}V_1 + (1 - \Lambda)(k - 1)(1 - F(y))^{k-2}F(y)V_2 \\ & -cx + \Lambda(1 - F(y))^{2k-2}V_1 + (\Lambda)(k - 1)^2(1 - F(y))^{2k-4}F^2(y)V_2 \\ & + 2\Lambda(k - 1)(1 - F(y))^{2k-3}F(y)V_2\big\}. \end{aligned} \tag{2.22}$$

Using the strict monotonicity of b and symmetry, the first order condition (FOC) is given by:

$$\begin{aligned}
& \left(-(1-\Lambda)(k-1)(1-F(y))^{k-2}F'(y)y' - \Lambda(2k-2)(1-F(y))^{2k-3}F'(y)y'\right) V_1 \frac{1}{y} \\
+ \quad & \left(-(1-\Lambda)(k-1)(1-F(y))^{k-3}F'(y)y'(1-(k-1))F(y)\right. \\
+ \quad & \left. 2\Lambda(k-1)(1-F(y))^{2k-5}F'(y)y'(1-kF(y)-((k-1)^2-1)F(y)^2)\right) V_2 \frac{1}{y} = 1
\end{aligned} \tag{2.23}$$

A contestant with the highest possible type $c = 1$ never wins a prize under the assumption $k > 2$. Thus the optimal effort of this contestant is always 0, providing $y(0) = 1$ as a boundary condition.

Note that the FOC is a differential equation with separated variables, since the left hand side of the equation (2.33) is a function of y only. Denote

$$\begin{aligned}
H(y) = \quad & V_1 \left((1-\Lambda)(k-1) \int_y^1 \frac{1}{t}(1-F(t))^{k-2}F'(t)dt + \Lambda(2k-2) \int_y^1 \frac{1}{t}(1-F(t))^{2k-3}F'(t)dt \right) \\
& + V_2 \left((1-\Lambda)(k-1) \int_y^1 \frac{1}{t}(1-F(t))^{k-3}(1-(k-1))F(t)F'(t)dt \right. \\
& \left. + 2\Lambda(k-1) \int_y^1 \frac{1}{t}(1-F(t))^{2k-5}F'(t)(1-kF(t)-((k-1)^2-1)F(t)^2)dt \right).
\end{aligned}$$

The solution to the differential equation (2.33) with the boundary condition $y(0) = 1$ becomes:

$$\int_x^0 dt = -H(y). \tag{2.24}$$

Equation (2.28) gives $x = H(y) = H(b^{-1}(x))$ implying $b = H$. In other words, the effort function of each player is given by $b(c) = A(c)V_1 + B(c)V_2$, where

$$\begin{aligned}
A(c) =& (1-\Lambda) \int_c^1 \frac{1}{a}(k-1)(1-F(a))^{k-2}F^{'}(a)da \\
& + \Lambda \int_c^1 \frac{1}{a}(2k-2)(1-F(a))^{2k-3}F^{'}(a)da
\end{aligned}$$

and

$$\begin{aligned}
B(c) =& (1-\Lambda) \int_c^1 \frac{1}{a}(k-1)(1-F(a))^{k-3}\left(-1+(k-1)F(a)\right)F^{'}(a)da \\
& + \Lambda \int_c^1 \frac{1}{a}(2k-2)(1-F(a))^{2k-5}\left(-1+kF(a)+((k-2)^2-1)F(a)^2\right)F^{'}(a)da.
\end{aligned}$$

Note that the terms multiplied by Λ and $(1-\Lambda)$ in $A(c)$ correspond to $-F_1'(a)$ and $-(F_1^2(a))'$, and in $B(c)$ correspond to $-F_2'(a)$ and $-\left((F_2^2(a))' + (2F_1(a)F_2(a))'\right)$

respectively, yielding

$$A(c) = (1-\Lambda)\int_c^1 -\frac{1}{a}F_1'(a)da + \Lambda\int_c^1 -\frac{1}{a}(F_1^2(a))'da \tag{2.25}$$

and

$$B(c) = (1-\Lambda)\int_c^1 -\frac{1}{a}F_2'(a)da + \Lambda\int_c^1 -\frac{1}{a}\left((F_2^2(a))' + (2F_1(a)F_2(a))'\right)da. \tag{2.26}$$

It remains to show that the equilibrium effort function $b(c)$ is differentiable and strictly decreasing. The former one is obvious. To show that the effort function is strictly decreasing, consider the derivatives of $A(c)$ and $B(c)$:

$$\begin{aligned} A'(c) =& (1-\Lambda) - \frac{1}{c}(k-1)(1-F(c))^{k-2}F^{'}(c) \\ & - \Lambda\frac{1}{c}(2k-2)(1-F(c))^{2k-3}F^{'}(c) < 0 \end{aligned}$$

and

$$\begin{aligned} B'(c) =& (1-\Lambda) - \frac{1}{c}(k-1)(1-F(c))^{k-3}\left(-1+(k-1)F(c)\right)F^{'}(c) \\ & + \Lambda - \frac{1}{c}(2k-2)(1-F(c))^{2k-5}\left(-1+kF(c)+((k-2)^2-1)F(c)^2\right)F^{'}(c). \end{aligned}$$

The derivative of the effort function $b(c)$ becomes:

$$\begin{aligned} b'(c) = \quad & A'(c)V_1 + B'(c)V_2 \\ & \leq V_2\left(A'(c)+B'(c)\right) \\ & < 0 \end{aligned}$$

since $V_2 \leq V_1$ and $B'(c)$ is smaller than $A'(c)$ in magnitude. Thus $b(c)$ is strictly decreasing.

Now suppose that contestants are sufficiently loss-averse, that is $\Lambda > 1$. In this case equation (2.31) implies that a non-negative expected pay-off from participating in the contest results for a contestant with type c only if

$$\frac{(F_1(c))^2V_1 + (F_2(c))^2V_2 + 2F_1(c)F_2(c)V_2}{F_1(c)V_1 + F_2(c)V_2} > 1 - \frac{1}{\Lambda}. \tag{2.27}$$

By Lemma 2.1 there exists a critical type, $\tilde{c}$, such that for all types $c < \tilde{c}$ equation (2.27) is satisfied while for all types $c > \tilde{c}$ it is violated. In order to secure a nonnegative pay-off all contestants with $c > \tilde{c}$ exert 0 effort in equilibrium. Note that $\tilde{c} = 1$ whenever $\Lambda \leq 1$.

The maximization problem of the agents remains the same, however the boundary condition becomes $y(0) = \tilde{c}$ when $\Lambda > 1$. Denote

$$\begin{aligned}\tilde{H}(y) = \quad & V_1\left((1-\Lambda)(k-1)\int_y^{\tilde{c}}\frac{1}{t}(1-F(t))^{k-2}F'(t)dt + \Lambda(2k-2)\int_y^{\tilde{c}}\frac{1}{t}(1-F(t))^{2k-3}F'(t)dt\right)\\ & +V_2\left((1-\Lambda)(k-1)\int_y^{\tilde{c}}\frac{1}{t}(1-F(t))^{k-3}(1-(k-1))F(t)F'(t)dt\right.\\ & \left.+2\Lambda(k-1)\int_y^{\tilde{c}}\frac{1}{t}(1-F(t))^{2k-5}F'(t)(1-kF(t)-((k-1)^2-1)F(t)^2)dt\right).\end{aligned}$$

The solution to the differential equation (2.33) with the new boundary condition becomes:

$$\int_x^0 dt = -\tilde{H}(y). \tag{2.28}$$

Using the same arguments as in the case of $\Lambda \leq 1$, the effort function of each contestant with type $c \leq \tilde{c}$ is given by $b(c) = A(c)V_1 + B(c)V_2$, where

$$\begin{aligned}A(c) =& (1-\Lambda)\int_c^{\tilde{c}}\frac{1}{a}(k-1)(1-F(a))^{k-2}F^{'}(a)da\\ &+\Lambda\int_c^{\tilde{c}}\frac{1}{a}(2k-2)(1-F(a))^{2k-3}F^{'}(a)da\end{aligned}$$

and

$$\begin{aligned}B(c) =& (1-\Lambda)\int_c^{\tilde{c}}\frac{1}{a}(k-1)(1-F(a))^{k-3}\left(-1+(k-1)F(a)\right)F^{'}(a)da\\ &+\Lambda\int_c^{\tilde{c}}\frac{1}{a}(2k-2)(1-F(a))^{2k-5}\left(-1+kF(a)+((k-2)^2-1)F(a)^2\right)F^{'}(a)da.\end{aligned}$$

Substituting $F_1(c)$ and $F_2(c)$, the weights of the first and the second prizes become:

$$A(c) = (1-\Lambda)\int_c^{\tilde{c}} -\frac{1}{a}F_1'(a)da + \Lambda\int_c^{\tilde{c}} -\frac{1}{a}(F_1^2(a))'da \tag{2.29}$$

and

$$B(c) = (1-\Lambda)\int_c^{\tilde{c}} -\frac{1}{a}F_2'(a)da + \Lambda\int_c^{\tilde{c}} -\frac{1}{a}\left((F_2^2(a))' + (2F_1(a)F_2(a))'\right)da. \tag{2.30}$$

Note that the weights of the first and the second prize are the same for any value of Λ, with the critical type being the least able contestant $\tilde{c} = 1$ whenever $\Lambda \leq 1$.

The optimal effort function is differentiable and strictly decreasing when $\tilde{c} < 1$, similar to the case of $\tilde{c} = 1$.

Q.E.D.

Proof of Proposition 2.3. The equilibrium effort function in the case of convex or concave cost-of-effort is derived in a similar to the case of linear cost-of-effort. As in the case of linear cost-of-effort, I will derive the optimal effort function first for the case when there is full participation in the contests (when $\Lambda \leq 1$) and then for the case when some contestants drop out the contest (when $\Lambda > 1$).

Assume that all contestants except i exert effort according to the function b which is strictly monotonic and differentiable. Suppose that each contestant participates in the contest, that is $\Lambda \leq 1$. The maximization problem of the contestant i with convex or concave cost-of-effort $\gamma(x)$ is:

$$\begin{aligned}\max x \quad & \{p_1\{V_1 + \eta(p_2(V_1 - V_2) + (1 - p_1 - p_2)V_1) - c\gamma(x)\} \\ & + \ p_2\{V_2 + \eta(p_1\lambda(V_2 - V_1) + (1 - p_1 - p_2)V_2) - c\gamma(x))\} \\ & + \ (1 - p_1 - p_2)\{\eta(p_1\lambda(-V_1) + p_2\lambda(-V_2)) - c\gamma(x)\}\} . \end{aligned} \tag{2.31}$$

where the probabilities of winning the first and the second prize ,p_1 and p_2, are defined as in equation (2.21). Denote the inverse effort function b^{-1} by y. Substituting b^{-1} and $\Lambda = \eta(\lambda - 1)$ and rearranging the terms, the maximization problem becomes:

$$\begin{aligned} max_x \quad & \left\{(1 - \Lambda)(1 - F(y))^{k-1}V_1 + (1 - \Lambda)(k - 1)(1 - F(y))^{k-2}F(y)V_2 \right. \\ & -c\gamma(x) + \Lambda(1 - F(y))^{2k-2}V_1 + (\Lambda)(k - 1)^2(1 - F(y))^{2k-4}F^2(y)V_2 \\ & \left. + \, 2\Lambda(k - 1)(1 - F(y))^{2k-3}F(y)V_2\right\} . \end{aligned} \tag{2.32}$$

.

Using the strict monotonicity of b and symmetry, the first order condition (FOC) is given by:

$$\begin{aligned} & \left(-(1 - \Lambda)(k - 1)(1 - F(y))^{k-2}F'(y)y' - \Lambda(2k - 2)(1 - F(y))^{2k-3}F'(y)y'\right) V_1\frac{1}{y} \\ + \quad & \left(-(1 - \Lambda)(k - 1)(1 - F(y))^{k-3}F'(y)y'(1 - (k - 1))F(y)\right. \\ + \quad & \left. 2\Lambda(k - 1)(1 - F(y))^{2k-5}F'(y)y'(1 - kF(y) - ((k - 1)^2 - 1)F(y)^2\right) V_2\frac{1}{y} = \gamma'(x) \end{aligned} \tag{2.33}$$

Using the boundary condition $y(0) = 1$, the solution to this differential equation is given by $\gamma(x) = H(y)$, where $H(y)$ is given by equation (2.24). Thus $x = \gamma^{-1}(H(y))$ implying that $b = \gamma^{-1}(H)$. The effort function of each contestant is given by $b(c) = \gamma^{-1}(A(c)V_1 + B(c)V_2)$, where $A(c)$ and $B(c)$ are given by equation (2.9) and (2.10) with

$\tilde{c} = 1$ respectively.

Now suppose that contestants are sufficiently loss-averse, that is $\Lambda > 1$. In this case by Lemma 2.1 there exists a critical type, $\tilde{c}$, such that for all types $c < \tilde{c}$ equation (2.27) is satisfied while for all types $c > \tilde{c}$ it is violated. Recall that contestants with $c > \tilde{c}$ exert 0 effort in equilibrium. Note that $\tilde{c} = 1$ whenever $\Lambda \leq 1$.

The maximization problem of the agents remains the same, however the boundary condition becomes $y(0) = \tilde{c}$ when $\Lambda > 1$. The solution to the differential equation (2.33) with the new boundary condition becomes $\gamma(x) = \tilde{H}(y)$, where $\tilde{H}(c)$ is given by equation (2.28). The effort function of each contestant is then given by $b(c) = \gamma^{-1}(A(c)V_1 + B(c)V_2)$, where $A(c)$ and $B(c)$ are given by equations (2.9) and (2.10) respectively.

It remains to show that the equilibrium effort function $b(c)$ is differentiable and strictly decreasing. The former one is obvious. To show that the effort function is strictly decreasing, consider the derivative of the effort function, $b'(c)$:

$$
\begin{aligned}
b'(c) = \quad & \gamma^{-1}(A(c)V_1 + B(c)V_2)\,(A'(c)V_1 + B'(c)V_2) \\
& < 0
\end{aligned}
$$

Using the proof of Proposition 1 and the fact that $\gamma^{-1} > 0$, one concludes that $b(c)$ is strictly decreasing.

Q.E.D.

2.B. Optimal Allocation of Prizes

Proof of Proposition 2.2. Assume that there are two prizes $V_1 \geq V_2 \geq 0$ to be awarded and $k > 2$ contestants. Assume that contestants have linear cost-of-effort functions. By Proposition 2.1 the average effort of each contestant is given by:

$$
\int_m^{\tilde{c}} b(c)F'(c)dc = \int_m^{\tilde{c}} (1-\alpha)A(c) + \alpha B(c)F'(c)dc. \tag{2.34}
$$

where $A(c)$ and $B(c)$ are given by equations (2.9) and (2.10). Note that $\tilde{c} = 1$ whenever $\Lambda \leq 1$. The designer's problem becomes:

$$
\max_{0 \leq \alpha \leq 1/2} k \int_m^{\tilde{c}} (A(c) + \alpha(B(c) - A(c)))F'(c)dc. \tag{2.35}
$$

Equivalently

$$\max_{0\leq\alpha\leq 1/2} \alpha \int_m^{\tilde{c}} (B(c) - A(c))F'(c)dc. \tag{2.36}$$

It is optimal to award a single first prize if and only if the integral in equation (2.36) is negative. Otherwise the optimal prize structure consists of two equal prizes, due to the linearity of the program. *Q.E.D.*

Proof of Proposition 2.4. Assume that there are two prizes $V_1 \geq V_2 \geq 0$ to be awarded and $k > 2$ contestants with either convex or concave cost-of-effort functions. By Proposition 2.3 the average effort of each contestant is given by:

$$\int_m^{\tilde{c}} b(c)F'(c)dc = \int_m^{\tilde{c}} \gamma^{-1} \left((1-\alpha)A(c) + \alpha B(c)\right) F'(c)dc. \tag{2.37}$$

where $A(c)$ and $B(c)$ are given by equations (2.9) and (2.10). Note that $\tilde{c} = 1$ whenever $\Lambda \leq 1$. The designer's revenue is given by:

$$R(\alpha) = k \int_m^{\tilde{c}} \gamma^{-1} \left((1-\alpha)A(c) + \alpha B(c)\right) F'(c)dc. \tag{2.38}$$

The designer's problem becomes:

$$\max_{0\leq\alpha\leq 1/2} k \int_m^{\tilde{c}} \gamma^{-1} \left((1-\alpha)A(c) + \alpha B(c)\right) F'(c)dc. \tag{2.39}$$

If condition in equation (2.19) is not satisfied, that is $R'(0) < 0$, then the integral in equation (2.39) is maximized at $\alpha = 0$. If, however, condition in equation (2.19) is satisfied, then $R(\alpha)$ can not have a maximum at $\alpha = 0$. It has a maximum at α^* with $R'(\alpha^*) = 0$.

Q.E.D.

Proof of Corollary 2.1. The revenue of the contest designer is given by:

$$R(\alpha) = k \int_m^1 \gamma^{-1}(A(c) + \alpha(B(c) - A(c)))F'(c)dc$$

Taking the second derivative of the revenue function with respect to α we get:

$$R''(\alpha) = k \int_m^1 \gamma^{-1''}(A(c) + \alpha(B(c) - A(c)))(B(c) - A(c))^2 F'(c)dc$$

Since the cost-of-effort function is concave, γ^{-1} is convex so that $\gamma^{-1''} > 0$. $((B(c) - A(c))^2 > 0$. Combining the two, $R''(\alpha) > 0$ implying that $R(\alpha)$ is convex in α. Therefore, the maximum of the revenue function is either at corner values, either $\alpha = 0$ or $\alpha =$

0.5. Note that $(B(c) - A(c))$ can not be zero, since there is always a positive mass of contestants exerting positive effort. Put differently, there exists a $\epsilon > 0$ such that contestants with abilities in $(m, m + \epsilon)$ exert positive effort.

Q.E.D.

2.C. The Symmetric Equilibrium with p Prizes

Assume that there are $2 < p \leq k$ prizes to be awarded with $V_1 \geq V_2 \geq \cdots \geq V_p$ and $k > p$ contestants. Assume that the cost-of-effort of contestants is given by $c\gamma(x)$, where γ is allowed to be linear, convex or concave. $F_s(a)$ denotes the probability that contestant i with type a meets $k-1$ competitors such that $s-1$ of these competitors have lower types than i and remaining $k-s$ competitors have higher types than i. $F_s(a)$ is then given by:

$$F_s(a) = \binom{k-1}{s-1}(1 - F(a))^{k-2}(F(a))^{s-1}$$

The expected utility of contestant i with cost parameter c is given by:

$$\begin{aligned}
EU =& \sum_{p=1}^{P} F_p V_p + \eta(\sum_{i>p} F_p F_i (V_p - V_i) + \sum_{i<p} F_p F_i \lambda (V_p - V_i)) - F_p c\gamma(x) \\
&+ \eta \sum_{p=1}^{P} (1 - \sum_{i=1}^{P} F_i) F_p \lambda (0 - V_p) \\
&+ (1 - \sum_{i=1}^{P} F_i) c\gamma(x).
\end{aligned}$$

Substituting $\Lambda = \eta(\lambda - 1)$ and rearranging the terms, the expected utility of contestant i becomes:

$$\begin{aligned}
EU =& \sum_{p=1}^{P} F_p V_p - c\gamma(x) \\
&- \Lambda \left\{ \sum_{i<p}^{P} F_p F_i (V_p - V_i) + \sum_{i=1}^{P} F_p (1 - \sum_{i=1}^{P}) V_p \right\}.
\end{aligned}$$

Rearranging the terms, one gets:

$$EU = (1 - \Lambda) \sum_{p=1}^{P} F_p V_p + \Lambda \left\{ \sum_{p=1}^{P} F_p^2 V_p + \sum_{i<p,p=2}^{P} 2 V_p F_p \sum_{i} F_1 \right\} - c\gamma(x).$$

The maximization problem of contestant i reads:

$$\max_x (1-\Lambda)\sum_{p=1}^{P} F_p V_p + \Lambda\left\{\sum_{p=1}^{P} F_p^2 V_p + \sum_{i<p,p=2}^{P} 2V_p F_p \sum_i F_1\right\} - c\gamma(x).$$

First order condition becomes:

$$\sum_{p=1}^{P} V_p \left\{(1-\Lambda)F_p' + \Lambda\left((F_p^2)' + \sum_{i<p}(2F_pF_i)'\right)\right\} = c\gamma^{-1\prime}(x)$$

The equilibrium effort function of contestant i with cost parameter c whose loss-aversion degree is smaller than 1, that is $\Lambda \leq 1$, becomes:

$$b(c) = \gamma^{-1}\left(\sum_{s}^{p} V_s \left\{(1-\Lambda)\int_c^1 -\frac{1}{a}F_s(a)'da \right.\right. \tag{2.40}$$
$$\left.\left. +\Lambda\int_c^1 -\frac{1}{a}\left(\left(F_s(a)^2\right)' + \sum_{i=1}^{s-1}\left(2F_i(a)F_s(a)\right)'\right) da\right\}\right)$$

Whenever $\Lambda > 1$, analogous to Lemma 2.1, there exist a critical type $\tilde{c}$ satisfying the following equation

$$\frac{\sum_{s=1}^{p}(F_s(\tilde{c}))^2 V_s + \sum_{s=2,i<s}^{p} 2V_sF_s(\tilde{c})F_i(\tilde{c})}{\sum_{s=1}^{p} F_s(\tilde{c})V_s} = 1 - \frac{1}{\Lambda}.$$

such that any contestant with $c \geq \tilde{c}$ exerts zero effort in equilibrium, while contestants with $c < \tilde{c}$ exert effort in equilibrium according to equation (40).

2.D. Allocation of p Prizes for Linear Costs

Assume that there are $2 < p \leq k$ prizes to be awarded with $V_1 \geq V_2 \geq ... \geq V_{p-1} \geq V_p \geq 0$ and $k > 2$ contestants. $F_s(a)$ denotes the probability that a contestant with type a wins the s-th prize, given by

$$F_s(a) = \binom{k-1}{s-1}(1-F(a))^{k-2}(F(a))^{s-1}.$$

If $\Lambda > 1$, then there exists a critical type $\tilde{c}$ such that in equilibrium each contestant with $c \geq \tilde{c}$ exerts 0 effort while each contestant with $c < \tilde{c}$ exerts effort according to

$$b(c) = \sum_{s}^{p} V_s \left\{ (1-\Lambda) \int_{c}^{\tilde{c}} -\frac{1}{a} F_s(a)' da \right.$$
$$\left. + \Lambda \int_{c}^{\tilde{c}} -\frac{1}{a} \left(\left(F_s(a)^2\right)' + \sum_{i=1}^{s-1} \left(2F_i(a)F_s(a)\right)' \right) da \right\}.$$

If $\Lambda \leq 1$, each contestant exerts effort according to the above equation with $\tilde{c} = 1$. Denote the coefficient of V_s by A_s:

$$A_s = \left\{ (1-\Lambda) \int_{c}^{\tilde{c}} -\frac{1}{a} F_s(a)' da \right.$$
$$\left. + \Lambda \int_{c}^{\tilde{c}} -\frac{1}{a} \left(\left(F_s(a)^2\right)' + \sum_{i=1}^{s-1} \left(2F_i(a)F_s(a)\right)' \right) da \right\}.$$

Substituting this into the bidding function we get:

$$\begin{aligned} b(c) &= \sum_{s=1}^{p} V_s A_s(c) \\ &= \left(1 - \sum_{i=1}^{p-1} V_{i+1}\right) A_1(c) + \sum_{i=2}^{p} V_i A_i \\ &= A_1 + \sum_{i=2}^{p} V_i \left(A_i(c) - A_1(c)\right) \end{aligned}$$

The desiger's problem becomes:

$$max_{0 \leq V_i \leq \frac{1}{i}} k \int_{m}^{\tilde{c}} \left\{ A_1(c) + \sum_{i=2}^{p} V_i \left(A_i(c) - A_1(c)\right) \right\} F'(c) dc$$

subject to the following $p-1$ conditions:

$$\begin{aligned} 1 - \sum_{i=1}^{p} V_i &\geqslant V_2 \\ V_2 &\geqslant V_3 \\ &\vdots \\ V_{p-1} &\geqslant V_p \end{aligned} \tag{2.41}$$

Since A_1 does not have a coefficient of type V_i, deleting A_1 would not harm. Since the summation is finite, it is allowed to interchange the integral and the summation signs.

Then the maximization problem reads:

$$max_{0\leq V_i\leq\frac{1}{i}}\sum_{i=2}^{p}\left\{V_i\int_m^{\tilde{c}}\left(A_i(c)-A_1(c)\right)F'(c)dc\right\}$$

subject to eqaution (2.42). It is optimal to award a single first prize if and only if each summand in the maximization problem is zero, that is

$$\int_m^{\tilde{c}}\left(A_i(c)-A_1(c)\right)F'(c)dc<0.$$

for each $i\in\{2,\dots,p\}$. Otherwise, it is optimal to award equal prizes only, due to the linearity of the program. That is the constraints in equation (2.42) will all bind. To see this, suppose to the contrary that there is an interior solution. WLOG assume that the interior solution is $(\sigma,\varsigma,\tau,0,\dots,0)\in[0,1]^p$, where $\sigma>\varsigma>\tau$ and $\sigma+\varsigma+\tau=1$. For the sake of easiness denote $G_i:=\int_m^{\tilde{c}}\left(A_i(c)-A_1(c)\right)F'(c)dc>0$. Since this allocation is optimal, and σ, ς, τ are all positive it means that G_4 is positive(otherwise it would be optimal to transfer the weight τ to σ and ς). Since $\tau>0$, G_2 should be greater than both G_3 and G_4 (otherwise it would be optimal to transfer the weight τ to σ and ς). But then τ should take the biggest value it could take, which is in this case $\frac{1}{3}$ (otherwise$(\sigma,\varsigma,\tau,0,\dots,0)$ would not be optimal). Applying the same reasoning to both σ and ς, we conclude that $\sigma=\varsigma=\tau=\frac{1}{3}$. In order to obtain the optimal prize allocation one needs to evaluate the objective function only on the boundary values, namely on the set $\{(1,0,\dots,0),(\frac{1}{2},\frac{1}{2},0,\dots,0),\dots,(\frac{1}{p},\frac{1}{p},\dots,\frac{1}{p})\}$ and take the allocation which gives the maximum value. It is optimal to award $2\leq r\leq p$ equal prices if and only if

$$r=\arg\max_{j\in 2,\dots,p}\frac{1}{j}\sum_{i=2}^{j}\left\{\int_m^{\tilde{c}}\left(A_i(c)-A_1(c)\right)F'(c)dc\right\}.$$

2.E. Allocation of p Prizes for Convex or Concave Costs

Assume that there are $2<p\leq k$ prizes to be awarded with $V_1\geq V_2\geq\ldots\geq V_{p-1}\geq V_p\geq 0$ and $k>2$ contestants. As before, $F_s(a)$ denotes the probability that a contestant with type a wins the s-th prize, given by

$$F_s(a)=\binom{k-1}{s-1}(1-F(a))^{k-2}(F(a))^{s-1}.$$

Whenever $\Lambda > 1$, there exists a critical type $\tilde{c}$ satisfying the following equation

$$\frac{\sum_{s=1}^{p}(F_s(\tilde{c}))^2 V_s + \sum_{s=2,i<s}^{p} 2V_s F_s(\tilde{c}) F_i(\tilde{c})}{\sum_{s=1}^{p} F_s(\tilde{c}) V_s} = 1 - \frac{1}{\Lambda}.$$

such that any contestant with $c \geq \tilde{c}$ exerts zero effort in equilibrium, while contestants with $c < \tilde{c}$ exert effort in equilibrium according to

$$\begin{aligned} b(c) = \gamma^{-1} \Bigg(\sum_{s}^{p} V_s \Bigg\{ (1-\Lambda) \int_c^{\tilde{c}} -\frac{1}{a} F_s(a)' da \\ +\Lambda \int_c^{\tilde{c}} -\frac{1}{a} \left(\left(F_s(a)^2\right)' + \sum_{i=1}^{s-1} \left(2F_i(a)F_s(a)\right)' \right) da \Bigg\} \Bigg) \end{aligned}$$

Note that $\tilde{c} = 1$ whenever $\Lambda \leq 1$. Denote the coefficient of V_s by A_s:

$$\begin{aligned} A_s = \Bigg\{ (1-\Lambda) \int_c^{\tilde{c}} -\frac{1}{a} F_s(a)' da \\ +\Lambda \int_c^{\tilde{c}} -\frac{1}{a} \left(\left(F_s(a)^2\right)' + \sum_{i=1}^{s-1} \left(2F_i(a)F_s(a)\right)' \right) da \Bigg\}. \end{aligned}$$

Substituting this into the optimal effort function we get:

$$\begin{aligned} b(c) =& \gamma^{-1} \left(\sum_{s=1}^{p} V_s A_s(c) \right) \\ =& \gamma^{-1} \left(\left(1 - \sum_{i=1}^{p-1} V_{i+1} \right) A_1(c) + \sum_{i=2}^{p} V_i A_i \right) \\ =& \gamma^{-1} \left(A_1 + \sum_{i=2}^{p} V_i \left(A_i(c) - A_1(c) \right) \right) \end{aligned}$$

The desiger's problem becomes:

$$max_{0 \leq V_i \leq \frac{1}{i}} k \int_m^{\tilde{c}} \gamma^{-1} \left(A_1(c) + \sum_{i=2}^{p} V_i \left(A_i(c) - A_1(c) \right) \right) F'(c) dc$$

subject to the following $p-1$ conditions:

$$\begin{aligned} 1 - \sum_{i=1}^{p} V_i &\geqslant V_2 \\ V_2 &\geqslant V_3 \\ &\vdots \\ V_{p-1} &\geqslant V_p \end{aligned} \tag{2.42}$$

So that the desiger's revenue is:

$$R(V_2, \cdots, V_p) = k \int_m^{\tilde{c}} \gamma^{-1} \left(A_1(c) + \sum_{i=2}^{p} V_i \left(A_i(c) - A_1(c) \right) \right) F'(c) dc$$

For the designer it is optimal to allocate the total prize sum into a single first one only if the partial derivatives $\frac{\partial R}{\partial V_h}(V_2, \cdots, V_p) \leq 0$ for each $h \in \{2, \cdots, P\}$. Thus, a sufficient for the optimality of multiple prizes is given by:

$$\int_m^{\tilde{c}} \gamma^{-1\prime} \left(A_1 + \sum_{i=2, i \neq h}^{P} V_i (A_i - A_1) \right) (A_h - A_1) dF'(c) dc > 0,$$

for each $h \in \{2, \cdots, P\}$.

Chapter 3

Cooperation in indefinitely repeated Games of Strategic Complements and Substitutes

3.1. Introduction

The study of cooperation and its determinants has attracted a great deal of attention in the literature. It is well-known, for instance, that in indefinitely repeated games, cooperation can be supported in equilibrium if the discount factor is sufficiently high (Friedman (1971)). Not much is known, however, about how the strategic environment—whether actions are strategic complements or substitutes—influences cooperative behavior in indefinitely repeated games. Strategic complementarity refers to the property that best-response functions are upward-sloping, whereas under strategic substitutability best-response functions are downward-sloping.[5] The distinction is relevant in several applications. For example, depending on whether firms in oligopolistic markets with homogeneous goods are engaged in price or quantity competition, actions are strategic complements or substitutes, and *vice versa* in markets with complementary goods. Also, depending on whether skills of members in teams are complementary or substitutable, efforts of team members are strategic complements or substitutes. Moreover, depending on whether the production of a public good is characterized by increasing or decreasing returns, contributions are strategic complements or substitutes. We study in an experiment whether in indefinitely repeated games strategic complements or substitutes are more conducive to cooperative behavior.

In our experiment, pairs of subjects play games with an indeterminate final period that feature either strategic complementarity or strategic substitutability (borrowed from Potters and Suetens, 2009), henceforth referred to as PS. Across the two treatments, several variables are kept constant, namely, the actions and payoffs in the Nash equilibrium of the stage game and in the symmetric joint-payoff maximum, the optimal defection

[5] A game is characterized by strategic complements (substitutes) if $\forall i, j$ and $i \neq j$: $\partial^2\pi/\partial x_i \partial x_j > 0$ (< 0), implying that the best-response functions are upward- (downward-) sloping (see Topkis (1978); Bulow et al. (1985); Fudenberg and Tirole (1984)).

payoff and the absolute value of the slope of the stage-game best-response function.[6] Subjects know that after each period the game proceeds to a next period with a continuation probability of 0.9. In order to allow for learning across games, subjects play at least 20 repeated games. After a repeated game ends, players are randomly re-matched to play another repeated game with the same continuation probability. The treatments are designed so that cooperation at the joint-payoff maximum can be sustained as a subgame-perfect Nash equilibrium. In particular, the treatments have the same critical discount factor above which such "full" cooperation is supported by e.g. a grim trigger strategy.

On average, choices in our experiment do not differ significantly between the strategic complements and substitutes treatments. However, the aggregate data mask two countervailing results that are in line with two distinct sets of studies. The first of these results is that the percentage of choices at the joint-payoff maximum is significantly higher under strategic substitutes than under strategic complements. This result fits well with the notion that strategic risk related to cooperation at the joint-payoff maximum is lower under substitutes than under complements. Recent theoretical and experimental studies on indefinitely repeated prisoner's dilemma games show that strategic risk is an important determinant of behavior. In particular, Blonski et al. (2011) formalize the intuition that cooperation gets riskier, and thus less likely, the more it hurts to cooperate if the partner defects (that is, the lower the "sucker" payoff). In particular, they propose a threshold for the discount factor in an indefinitely repeated game above which cooperation at the joint-payoff maximum is supported in equilibrium, which is higher than the standard threshold based on e.g. grim-trigger strategies (see also Blonski and Spagnolo (ress)). Blonski et al. (2011) and Dal Bó and Fréchette (2011) provide experimental evidence showing that this threshold is necessary for cooperation in a prisoner's dilemma to increase to very high levels.

The second result in our experiment is that choices of subjects in pairs that do not succeed in cooperating at the joint-payoff maximum tend to be lower, so less cooperative, under strategic substitutes than under strategic complements. This finding squares well with theoretical and experimental findings on the differential effects of strategic substitutes and complements on cooperation in the presence of heterogeneous player types. Indeed, if a cooperator is matched with a best-responder, the aggregate outcome in a pair will be less cooperative under strategic substitutes than under complements (Haltiwanger and Waldman, 1991, 1993; Camerer and Fehr, 2006). The reason is that a

[6]To be precise there are three differences between our setting and PS. While they have a finite game with 30 periods with fixed matching, we have repeated supergames with random matching across supergames. Moreover, while PS implement games with positive and negative externalities (with no reported significant difference between the two) we only implement games with positive externality.

best-response to a cooperative choice is less cooperative under strategic substitutes than under complements in the sense that it deviates less from the static Nash equilibrium in the former than in the latter case.[7] Experimental evidence for this intuition in the context of a "long" finitely repeated dilemma game is provided by PS.[8]

The remainder of this paper is organized as follows. In Section 2 we introduce the experimental design and procedures. In Section 3 we develop the conjectures concerning predicted behavior in our experiment, focusing on the comparative static predictions between the treatments with complements and substitutes. In Section 4 we present the experimental results. In Section 5 we summarize and discuss our findings in the light of the existing literature.

3.2. Experimental Design and Procedures

3.2.1. Experimental Design

Our experiment has two treatments: one where choices are strategic complements (Comp) and another where choices are strategic substitutes (Subs). In each treatment, subjects play an indefinite repetition of the same stage game. The stage game has a unique and Pareto dominated Nash equilibrium and a symmetric socially efficient (joint payoff maximizing) outcome (JPM). The payoffs in each treatment are determined according to the following payoff functions (borrowed from PS):

$$\pi_i^{\mathsf{Comp}}(x_i, x_j) = -28 + 5.474x_i + 0.01x_j - 0.278x_i^2 + 0.0055x_j^2 + 0.165x_ix_j, \tag{3.1}$$

$$\pi_i^{\mathsf{Subs}}(x_i, x_j) = -28 + 2.969x_i + 2.515x_j - 0.082x_i^2 + 0.023x_j^2 - 0.0485x_ix_j. \tag{3.2}$$

The coefficients in the payoff functions are chosen in order to ensure a fair comparison between the two treatments. First, in both treatments, the stage game has the same Nash equilibrium and the same JPM outcome. That is, the positions of the Nash equilibrium and the JPM choice are the same in the two treatments. Second, the payoffs corresponding to the Nash equilibrium and the JPM are the same across the two treatments. Third, the payoff achieved by best responding to JPM play of the matched player, referred to

[7] Note that, in contrast to theory, the estimated response functions in the two treatments have the same positive sign, but it still holds that the slope in the case of complements is larger than the one in the case of substitutes.

[8] See Haltiwanger and Waldman (1985) and Fehr and Tyran (2008) for applications where aggregate outcomes depend on the strategic environment if individuals are heterogeneous in the rationality of their expectations.

as the defection payoff, is the same in the two treatments.[9] Lastly, the absolute value of the slopes of the best-response curves are the same in the two treatments to guarantee that the same speed of convergence is generated by the best-response dynamics. Table 4.1 summarizes the main theoretical benchmarks of our design.

Table 3.1: Theoretical Benchmarks

	Comp	Subs
$Choice_{Nash}$	14.0	14.0
$Choice_{JPM}$	25.5	25.5
Π_{Nash}	27.71	27.71
Π_{JPM}	41.97	41.97
Π_{Defect}	60.14	60.14
$Slope\ of\ reaction\ function$	0.30	-0.30

Notes: This table shows the theoretical benchmarks regarding choices and payoffs in the experiment.

In order to allow for learning, in our experiment subjects played a series of one of the two games described above. We refer to each repeated game, that is, each sequence of periods determined by the continuation probability of 0.9, as a match. Once a match ended and depending on the time left, another one started. In each session, subjects participated in as many matches as possible such that at least 20 matches were played. If at least 20 matches had already been played, a session ended after one and a half hours of play. Subjects played with the same partner throughout a match. Once a match ended, subjects were randomly re-matched with another subject.

By using the payoff functions given in (3.1) and (3.2), we keep several actions and payoffs constant across treatments. We felt the same should be done with respect to the sequence of matches and their respective lengths. At the same time, because of possible order effects, we did not only want to have one sequence of matches to be played in each of the two treatments. We therefore decided to have five different draws of the lengths of matches prior to the start of the experiments, each of which was administered in one session for each of the two treatments Comp and Subs.[10] The length of each match in a draw was determined randomly with the continuation probability of 0.9. Figure 3.15 in the Appendix shows the distribution of realized match lengths across all five draws.[11]

[9] The combination of the second and third condition mentioned above has as the consequence that payoffs on the best-response function are the same in the two treatments.

[10] For instance, under draw number 1, the randomly determined lengths of the matches played was: 11, 5, 9, 5, 18, 33, 7, 7, 5, 12, 4, 16, 11, 1, 5, 4, 23, 9, 14, 6, 6, 10, 2, 7, 1.

[11] In an indefinitely repeated game with continuation probability $\delta = 0.9$, the expected number of periods in each match is 10.

Since there is always the possibility of continuing to a next round, the randomization generates a game that is strategically equivalent to an indefinitely repeated game. In particular, the continuation probability δ is equivalent to the discount factor in an indefinitely repeated game assuming that within the time slot of an experiment, there is no discounting (Roth and Murnighan (1978)).

3.2.2. Experimental Procedures

The experiment consists of 10 sessions (five for each of the two treatments Comp and Subs) that were conducted at CentERlab at Tilburg University during September-October 2011.[12] A total number of 160 students participated in the experiment. Participants were recruited through an email list of students who are interested in participating in the experiments. In each session, 16 subjects interacted anonymously in a sequence of matches, that is, indefinite repetitions of the same stage game. In each session subjects participated in between 20 and 25 matches. Each session lasted not more than two hours (including the time to read the instructions and payment of the subjects).

All participants were given the same instructions (see Appendix A). At the beginning of each match, subjects were randomly paired with each other. During a match, subjects played with the same partner. The matching rule was explained clearly before the experiment started. The identity of the partners was not revealed to subjects. It was explained to the subjects that their final earnings depended on their own choices and the choice of the matched participants. The subjects were asked to choose a number between 0.0 and 28.0 (up to one digit after the comma) in each round of a match. Subjects were provided an earnings calculator on the computer screen enabling them to calculate their earnings in points for any combination of hypothetical choices, and a payoff table for combinations of hypothetical choices that are multiples of two (see Table 3.7 and Table 3.8).

After choices were submitted in each round, subjects were informed about whether the match would continue to a next round or not. In the case the game continued to a next round, subjects received the message "*The match continues to the next round.*" on the computer screen. In the case the match ended, subjects received the message "*The match is over.*" on the computer screen. Once a match ended, another match would begin, depending on the time available. Moreover, after each round of a match the subjects were provided with information of the previous round on the screen, namely their own choice and earnings and the matched partner's choice and earnings.

[12] We used the experimental software toolkit *Z-Tree* to program and conduct the experiment (see Fischbacher (2007a).)

After the subjects finished reading the instructions, we explained to them that the experiment itself would proceed for about 1.5 hours.

The payoffs in the experiment was expressed in points. At the end of the experiment, the sum of a subject's earnings in points in all rounds of all matches were converted into Euro at the exchange rate of 480 points = 1 Euro, and privately paid to subjects. The average earnings in the experiment was 16.45 Euro.

3.3. Conjectures

The first conjecture builds on the predictions of the standard theory of indefinitely repeated games. Based on simple grim-trigger strategies, this theory predicts that cooperation can be supported as a subgame perfect Nash equilibrium (SPNE) if the following condition holds:

$$\frac{\Pi_{JPM}}{1-\delta} \geq \Pi_{Defect} + \frac{\delta \Pi_{Nash}}{1-\delta}. \tag{3.3}$$

The left-hand side of (3.3) is the discounted sum of payoffs from cooperation, while the right-hand side is the discounted sum of payoffs from a one-time deviation followed by Nash equilibrium play forever after. By design, the JPM payoff, the defection payoff, and the static Nash equilibrium payoff are the same in both treatments. Rearranging condition (3.3) and using the numbers given in Table 4.1, we get

$$\delta \geq \underline{\delta} := \frac{\Pi_{Defect} - \Pi_{JPM}}{\Pi_{Defect} - \Pi_{Nash}} = \frac{60.14 - 41.94}{60.14 - 27.71} = 0.56 \tag{3.4}$$

for both treatments. We thus conclude that the critical discount factor above which cooperation at the joint-payoff maximum (full cooperation) is supported by a grim-trigger strategy is the same in both treatments.[13], [14] This leads to our first conjecture.

SPNE Conjecture. *The full cooperation rate should be the same in* Subs *and* Comp*.*

The second conjecture takes into account differences in relative riskiness of cooperation between the two treatments. Inspecting the payoffs in Subs and Comp, one notices that if one player plays fully cooperatively, while the other player in the market defects

[13] The range of actions that Pareto-dominate the static Nash equilibrium, and thus also the range of actions that can be sustained in equilibrium in an indefinitely repeated game, is larger under substitutes than under complements. This can be seen in Figure 3.10 in the Appendix that shows the iso-payoff contours in both cases. Given the findings of Gazzale (2009), we did not expect that this difference would lead to differences in the extent to which subjects succeed in fully cooperating. It may lead to larger variability in actions under substitutes than under complements, though.

[14] However, note the following. Any feasible and admissible average payoff vector above the NE of the stage game can be supported as a SPNE provided that δ is sufficiently high. This area of this region for Comp is 386.648, while for Subs it is 403.246.

Table 3.2: A general and reduced PD games for the two treatments

(a) A general PD game

	C	D
C	c, c	a, b
D	b, a	d, d

(b) The reduced PD game for Comp

	C	D
C	41.94, 41.94	5.89, 60.14
D	60.14, 5.89	34.90, 34.90

(c) The reduced PD game for Subs

	C	D
C	41.94, 41.94	10.71, 60.14
D	60.14, 10.71	18.17, 18.17

Notes: This table illustrates the payoff matrices for a general PD game and the reduced PD games for Comp and Subs treatments.

optimally, the cooperating player's ("sucker") payoff is lower with complements than with substitutes. In addition, the payoff players get if they both optimally defect, is lower in Subs than in Comp. Intuitively, these two forces make it less attractive, because relatively more risky, to choose actions that maximize joint payoffs in Comp than in Subs.

Recently, this intuitive idea received formal support in Blonski et al. (2011). These authors suggest an axiomatic approach to equilibrium selection in indefinitely repeated prisoner's dilemma (PD) games. They show that a set of five axioms leads to a discount factor δ^* that is strictly larger than the standard discount factor $\underline{\delta}$ in the SPNE Conjecture and that, more importantly for our purposes, reflects the influence of the sucker payoff on the incidence of fully cooperative play.[15] In particular, given a PD stage game of the form shown in Panel (a) in Table 3.2 with $b > c > d > a$ and $2c > b + a$, Blonski, Ockenfels and Spagnolo (2011, Proposition 2) show that their five axioms imply the threshold $\delta^* = (b - c + d - a)/(b - a)$ above which a cooperation equilibrium is predicted to be played in the indefinitely repeated PD. Note that this threshold features the sucker payoff a, while the threshold of the SPNE Conjecture does not (there $\underline{\delta} = (b - c)/(b - d)$). Note also that $\partial\delta^*/\partial a = -(c - d)\,/\,(a - b)^2 < 0$, so that a *lower* sucker payoff *increases* the threshold above which cooperation should be observed. Put differently, the lower the sucker payoff, the smaller the range of discount factors for which a cooperation equilibrium is selected.

Blonski et al. (2011) develop their approach in the context of a standard 2×2 PD game. Our stage game, however, has many more than just two actions. Still, we believe that the intuitive idea that a lower "sucker" payoff and higher "mutual optimal defection" payoff should *ceteris paribus* lead to less full cooperation is also relevant in the context of our stage games. A conjecture that translates Blonski et al's approach to our context

[15] The five axioms in Blonski et al. (2011) are called (1) positive linear payoff transformation invariance; (2) δ-monotonicity, (3) boundary conditions (which is the crucial axiom that highlights the influence of the sucker payoff on the incidence of cooperation); (4) incentive independence; and (5) equal weight.

can be generated if one is willing to make the simplifying assumption that the action space of our stage games consist of just two choices, say Choice $C = Choice_{JPM}$ and $D = Choice_{Defect}$. Using the payoff functions given in (3.1) and (3.2), these two choices lead to the two games shown in Panels (b) and (c) in Table 3.2.[16] It follows that $\delta^*_{\mathsf{Comp}} = 0.870$ and $\delta^*_{\mathsf{Subs}} = 0.518$, so that full cooperation can be sustained for a larger range of discount factors in treatment Subs than in treatment Comp.[17]

An alternative concept leading to the same comparative static prediction is the *basin of attraction* of a cooperative strategy in comparison to a defecting strategy (see Dal Bó and Fréchette, 2011). To understand the idea of the basin of attraction, assume (again, a strong assumption) that players either play "tit for tat" (a cooperative strategy) or "always defect" (a defective strategy) and nothing else in the repeated PD game and that this is common knowledge. Then a player needs to determine which of these two strategies generates the higher expected payoff given the belief that with probability p the other player plays "tit for tat" and with probability $1-p$ plays "always defect". The basin of attraction of the cooperative strategy is the set of beliefs p for which playing this strategy gives a higher expected payoff than the defective strategy. In the context of the general game shown in Panel (a) in Table 3.2, the expected payoff for the cooperative strategy is equal to

$$\begin{aligned} C(a,c,d,\delta) &= p(c+\delta c+\delta^2 c+...)+(1-p)(a+\delta d+\delta^2 d+...) \\ &= 1/(1-\delta)\,(a-a\delta+d\delta-ap+cp+ap\delta-dp\delta)\,, \end{aligned} \tag{3.5}$$

while the expected payoff for the defecting strategy is equal to

$$\begin{aligned} D(a,b,d,\delta) &= p(b+\delta d+\delta^2 d+...)+(1-p)(d+\delta d+\delta^2 d+...) \\ &= 1/(1-\delta)\,(d+bp-dp-bp\delta+dp\delta)\,. \end{aligned} \tag{3.6}$$

Equating the two expressions in (3.5) and (3.6) gives the threshold p^* above which playing the cooperating strategy is the payoff maximizing choice. That is, the lower p^* the larger the basin of attraction of the cooperative strategy and the more likely it is that subjects will choose to fully cooperate. For the games shown in Panel (b) and (c) in Table 3.2, we find $p^*_{\mathsf{Comp}} = 0.391$ and $p^*_{\mathsf{Subs}} = 0.038$, so that, again, full cooperation is predicted to emerge for a larger range of beliefs in Subs than in Comp.[18] The lines of reasoning based

[16] The choice $C = Choice_{JPM}$ is equal to 25.5 in both treatments, while $D = Choice_{Defect} = 17.42$ in Comp and $D = Choice_{Defect} = 10.64$ in Subs.

[17] In case the two choices are $C = Choice_{JPM} = 25.5$ and $D = Choice_{Nash} = 14$ in both treatments, we find again $\delta^*_{\mathsf{Comp}} = 0.7834$ and $\delta^*_{\mathsf{Subs}} = 0.664$, and so, again, $\delta^*_{\mathsf{Comp}} > \delta^*_{\mathsf{Subs}}$.

[18] If the 2×2 PD games are generated using the actions mentioned in footnote 17, we find $p^*_{\mathsf{Comp}} = 0.784$

on the influence of the sucker payoff and the basin of attraction lead to:

Riskiness Conjecture. *The full cooperation rate should be higher in* Subs *than in* Comp.

The third conjecture is based on the literature that studies the interaction between the strategic environment (complements versus substitutes) and heterogeneity of players (Haltiwanger and Waldman (1991), Camerer and Fehr (2006)), as well as its application to repeated-game experiments (see PS). The intuition goes as follows. In games of strategic complements a change in the matched player's choice gives a payoff-maximizing player an incentive to move in the *same* direction, while in games of strategic substitutes the incentive is to move in the *opposite* direction. Given that several experiments have shown that some people are (conditionally) cooperative in the sense that they try to induce cooperation and follow it when established by others, even when there is no future interaction, (see Fehr and Fischbacher (2002), Clark and Sefton (2001) and Reuben and Suetens (2009)), it is plausible to assume that players are heterogeneous in their cooperativeness and defection strategies. Consider, for example, a cooperative player who is matched with a defector in the above-described games of complements and substitutes. If the cooperative player makes a cooperative choice (higher than the static Nash equilibrium), and the matched defector is an optimal defector in the sense that he best-responds to this move, then, in sum, choices will be higher (more cooperative) in Comp than in Subs. This is because in Comp, the best-response to a cooperative move is to (partly) follow the move and make a higher choice as well, whereas in Subs, the best-response is to make a less cooperative choice. This mechanism may facilitate cooperation in Comp and may hamper it in Subs. In addition, a similar mechanism occurs when a cooperative player is matched with a spiteful defector who aims at maximizing the payoff difference between himself and the cooperator. In order to employ the same level of punishment (in payoff terms), a spiteful defector must choose much lower choices in the Subs treatment than in the Comp treatment. So here as well, choices will, on average, be higher, e.g. more cooperative in Comp than in Subs. PS provide evidence for this intuition in the context of a finitely repeated game. We summarize this intuition in the following conjecture.

Heterogeneity Conjecture. *In pairs that do not succeed in joint full cooperation, choices should be higher (i.e. more cooperative) in* Comp *than in* Subs.

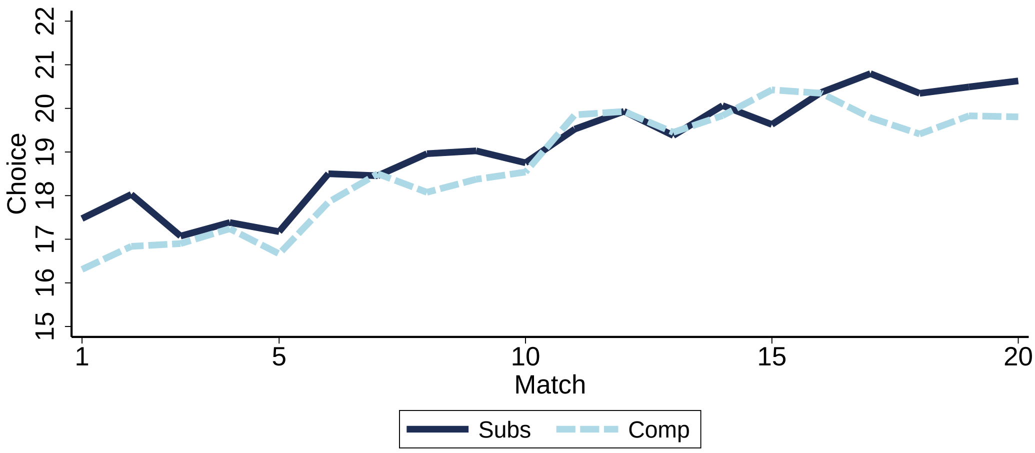
and $p^*_{\mathsf{Subs}} = 0.664$, and so again $p^*_{\mathsf{Comp}} > p^*_{\mathsf{Subs}}$.

3.4. Experimental Results

In this section we describe our main results. We analyze data from matches 1-20 for which we have observations in all sessions.

Averaged over all subjects, rounds and matches, the mean choice is 19.09 in treatment Subs and 18.70 in treatment Comp. In the last 10 matches the mean choice in the Subs treatment is 20.12 and that in the Comp treatment is 19.87.[19] The average choice is thus roughly the same in the two treatments.

Figure 3.1: Evolution of Average Choices



Notes: This figure shows the evolution of individual choices across matches for the treatments.

Figure 3.1 illustrates the evolution of average choices over time under strategic complements and strategic substitutes. In both treatments, the average choice is increasing over the matches. However, there is no clear difference between the two treatments.[20] To formally quantify the difference between the two treatments, and to test whether it is statistically significant, we estimate the effect of strategic complementarity on the individual choice. We do so by regressing the choice of an individual on a treatment dummy, and clustering standard errors at the session level (results are reported in column (1) of Table 3.5). The regression results confirm that the difference between the two treatments is small in size, and not statistically significant (marginal effect is -0.365, $p = 0.679$).[21,22]

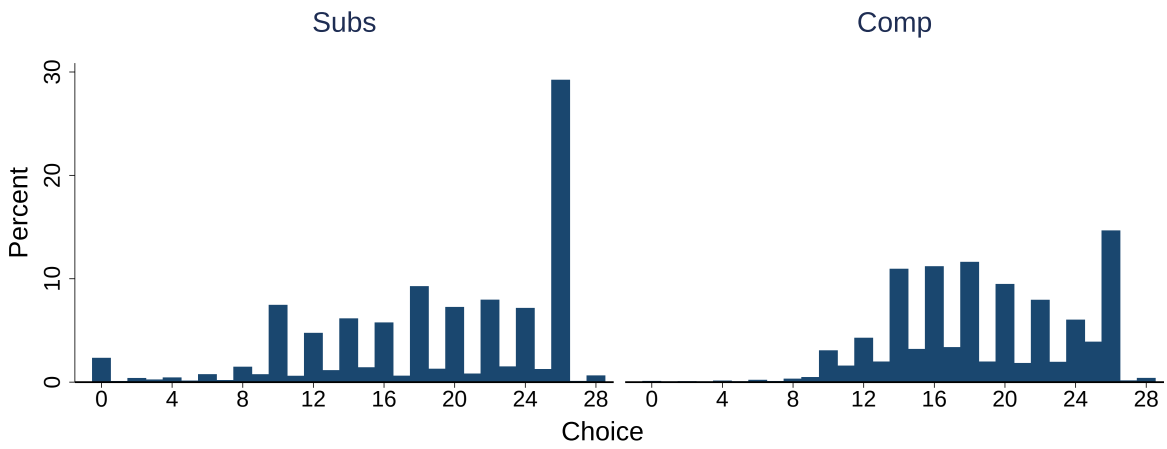
[19] The summary statistics for average choice is presented in Table 3.10 in the Appendix.

[20] There is also no significant difference in payoffs between the two treatments as reported in Table 3.9 in the Appendix.

[21] The estimated marginal effect of strategic complementarity on individual choice becomes -0.386 with $p = 0.505$ when we control for the match and the interaction between treatment and match. No significant differences are obtained in payoffs either. This can be seen in column (2) of Table 3.5.

[22] Mann-Whitney-U tests based on independent observations yield similar results, both when the

Figure 3.2: Distribution of Choices



Notes: This figure shows the distribution of individual choices in the experiment.

Some properties of the data might be hidden when looking at aggregates. To analyze the data in more detail, in a next step, we present the distribution of choices for strategic substitutes and complements. Figure 3.2 shows that choices in the Subs treatment are spread over the whole interval, while choices in the Comp treatment are somewhat more concentrated. Moreover, the modal choice in both treatments is a choice at or very close to the JPM level of 25.5. This is particularly accentuated in treatment Subs. To illustrate, in Subs, almost 30% of the choices are at or very close to the JPM level of 25.5, whereas in Comp we observe about 15% of such choices.

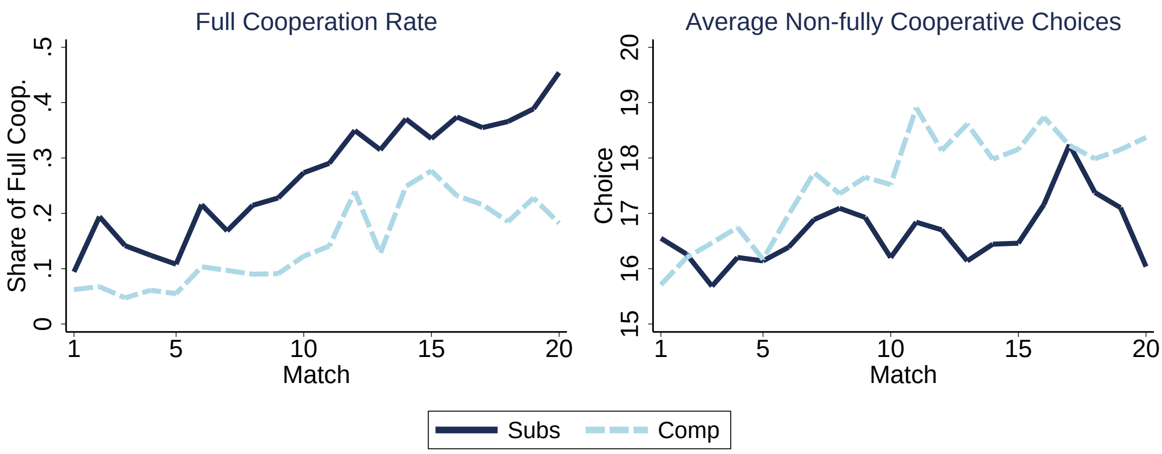
To further explore potential differences between Subs and Comp, we distinguish "fully-cooperative" and "non-fully cooperative" choices. We define a choice to be fully-cooperative if it lies within the interval $[25, 26]$, where 25.5 is the JPM choice in both treatments. We refer to a choice as non-fully cooperative if it lies outside the range $[25, 26]$.[23]

The left-hand panel of Figure 3.3 illustrates for both treatments the share of fully cooperative choices by match (so over time). From this graph it becomes clear that the share of fully cooperative choices is higher in Subs than in Comp. In addition, the share of fully cooperative choices increases in both treatments, but more so in Subs than in Comp. In the last 10 matches, the percentage of fully cooperative choices is around 40%

average choice is based on all matches, or the last 10 matches ($p = 0.750$ in both cases).

[23] The choice of such a range is to some extent arbitrary, and one may argue that choices above 26 are also fully cooperative. For example, 28, which is the maximum choice possible, can serve as a focal point for subjects to coordinate on (almost) full cooperation. Enlarging the fully-cooperative interval to $[25, 28]$, does not affect any of our qualitative results in what follows. Choices above 26 correspond to 0.68 % of all choices in the experiment.

Figure 3.3: Cooperative vs Non-Cooperative Behavior



Notes: This figure shows the evolution of cooperative and non-cooperative behavior. The left-hand panel depicts the evolution of full cooperation rate across matches and the right-hand panel depicts the evolution of average non-fully cooperative choices across matches.

in Subs, while it is around 25% in Comp.[24]

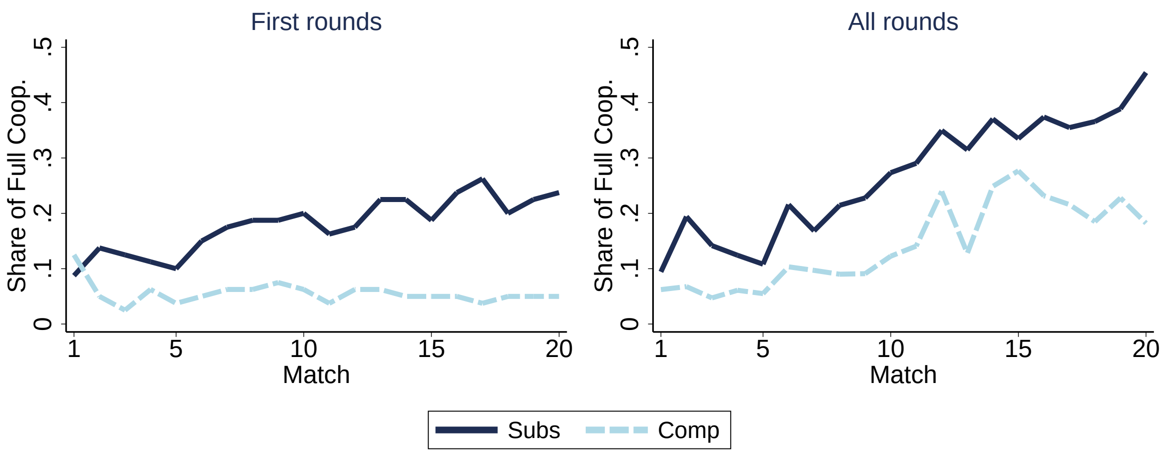
The right-hand panel of Figure 3.3 depicts the evolution over matches of averages of non-fully cooperative choices (those that fall outside the interval $[25, 26]$). Here we observe that the average choice of subjects is, overall, higher in Comp than in Subs. So it seems the effect of strategic complementarity on behavior switches—behavior is more cooperative because choices are higher—when we focus on non-fully-cooperative choices. To illustrate, averaged over subjects, rounds and matches, the mean non-fully cooperative choice is 16.65 in the Subs treatment and it is 17.59 in the Comp treatment. In the second half of the experiment, the average non-fully cooperative choice is 16.85 in Subs and 18.33 in Comp.[25]

In sum, although we do not observe a clear difference between the two treatments at the aggregate level, analyzing fully cooperative and non-fully cooperative behavior separately suggests that, overall, behavior is driven by two countervailing forces. On the one hand, subjects make choices at the fully cooperative level more frequently under Subs than under Comp. On the other hand, the average choice of subjects who do not make fully-cooperative choices is higher under Comp than under Subs. To understand which forces drive these two results, we analyze fully cooperative behavior in section 3.4.1 and non-fully cooperative behavior in 3.4.2 in more detail. Secion 3.4.3 focuses on results at the pair level.

[24]For an in-depth analysis of the statistical significance of these observations, see Section 3.4.1.

[25]For an in-depth analysis of the statistical significant, see Section 3.4.2.

Figure 3.4: Full Cooperation Rate



Notes: This figure shows the evolution of full cooperation rate across matches, on the left-hand panel for the first rounds only and on the right-hand panel for all rounds.

3.4.1. Full Cooperation Rates

In this section we take a closer look at full cooperation rates, that is, choices in the interval $[25, 26]$ at the level of subjects. In doing so, we examine the first and all rounds of a match separately since the cooperation rate might evolve within a match, depending on the number of rounds in that match (see, Dal Bó and Fréchette (2011)). In addition, in the first rounds of each match subjects are playing with a new partner so that they do not have experience with their partners' behavior. In this respect, subjects' behavior in the first round of each match is mainly driven by the fundamentals of the game they are playing (and possibly their experiences in the previous matches) and not by the partners' behavior.

Figure 3.4 illustrates the evolution of the full cooperation rate across matches, in the left-hand panel for the first rounds and in the right-hand panel for all rounds. The left-hand panel shows that in the first rounds of a match subjects make fully cooperative choices more frequently under Subs than under Comp. In addition, the first-round full cooperation rate follows an increasing trend in Subs, while in Comp it is more steady across matches. The consequence is that the full cooperation rate in the first match is almost the same for the two treatments, while towards the end of the experiment there is a considerable difference in full cooperation rates between the two treatments. The first-round full cooperation rate reaches the level of about 25% in the Subs treatment by the end of the experiment while it remains around 5% in the Comp treatment.

In order to test whether these differences are statistically significant, we ran two specifications of probit regressions where the dependent variable is a dummy referring to a subject fully cooperating or not. In the first specification shown in Table 3.3 we

Table 3.3: Regression results on full cooperation

	First rounds		All rounds	
	(1)	(2)	(3)	(4)
VARIABLES	FullCoop_{it}	FullCoop_{it}	FullCoop_{it}	FullCoop_{it}
Comp	−0.127***	−0.050***	−0.115***	−0.178***
	(0.031)	(0.022)	(0.042)	(0.043)
Round				0.004***
				(0.002)
Comp*Round				0.004**
				(0.002)
Match		0.005***		0.016***
		(0.002)		(0.002)
Comp*Match		−0.007***		0.001
		(0.003)		(0.004)
Observations	3,200	3,200	33,024	33,024

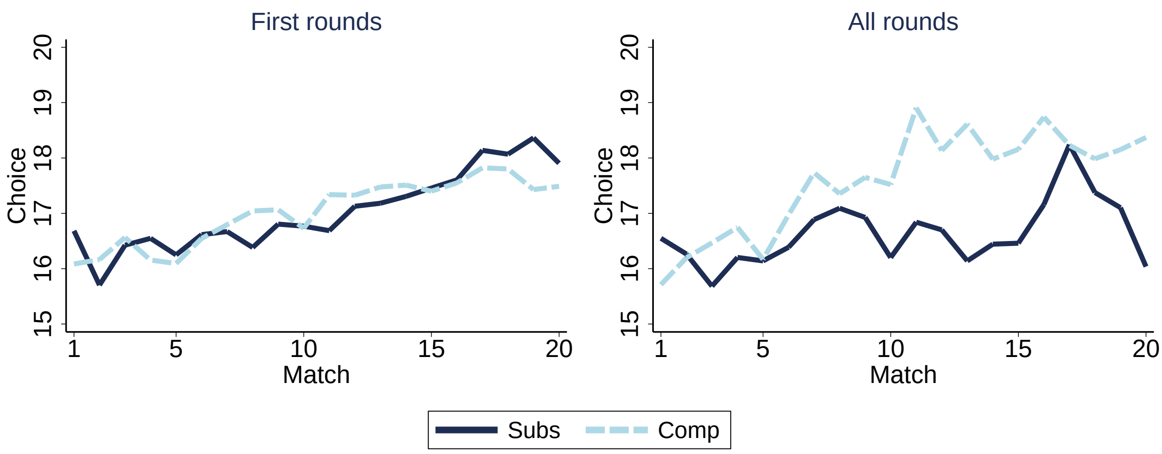
Notes: This table reports marginal effects from probit regressions with delta-method standard errors (in parentheses) clustered at the session level. The dependent variable is a dummy which is equal to 1 if the choice is fully cooperative and 0 otherwise. *** (**) [*] indicate that the estimated coefficient is significant at the 1% (5%) [10%] level. Specifications (1) and (2) are based on observations from the first rounds of matches only and specifications (3) and (4) are based on all observations.

include as an independent variable a treatment dummy. In the second specification, next to the treatment dummy, we control for the match, and the interaction between treatment and match. As shown in Table 3.3, in both specifications the treatment dummy has a negative sign—the full cooperation rate in Subs is thus lower than the one in Comp—and is statistically significant. The estimated marginal effect is −0.127 and −0.50 respectively.[26] In addition, column (2) shows that the first-round full cooperation rate significantly increases over the matches in Subs (marginal effect is 0.005, $p \leq 0.001$), but not so in Comp (marginal effect is -0.002, $p \leq 0.001$).

Next, we focus on the right-hand panel of Figure 3.4 and the remainder of Table 3.3. As illustrated in the figure, there is again a clear difference between the two treatments in the full cooperation rate. In contrast to the first rounds, the full cooperation rate now increases over matches in Comp as well. The full cooperation rate raises up to about

[26] The p-values in Mann-Whitney-U tests based on sessions averages are 0.016 if all matches are taken into account and 0.075 if only matches 11-20 are taken into account.

Figure 3.5: Average Non-Fully Cooperative Choices



Notes: This figure shows the evolution of non-fully cooperative choices (i.e. choices outside the range $[25, 26]$) across matches, on the left-hand panel for the first rounds only and on the right-hand panel for all rounds.

25% in Comp and up to about 45% in Subs.

The results of probit regressions, which we report in columns (3) and (4) of Table 3.3, indicate that the treatment effects are again statistically significant. Moreover, as shown in column (4), the full cooperation rate increases significantly over the matches in both treatments (marginal effect is 0.016, $p = 0.001$).

Summarizing, we find significantly more initiation of full cooperation at the beginning of a new match as well as more fully cooperative choices in general in Subs than in Comp. This result is in line with the "riskiness-of-cooperation" conjecture.

3.4.2. Non-Fully Cooperative Behavior

In this section we analyze the effect of strategic complementarity on non-fully-cooperative behavior. In doing so, we focus on those data points that are not in the fully cooperative range of $[25, 26]$. Figure 3.5 depicts the evolution of the average non-fully-cooperative choice over matches, in the left-hand panel for the first rounds and in the right-hand panel for all rounds of a match.

The figure in the left-hand panel shows that in the first rounds of the matches there is no visible difference in non-fully-cooperative behavior between the two treatments. The figure also shows that in both treatments the average non-fully cooperative choice in the first rounds is initially above the static Nash equilibrium choice of 14 and increases over the matches. As shown in Table 3.4 presenting results from linear regressions where the average non-fully-cooperative choice is regressed on a treatment dummy, the treatment effect is small and not significant. In addition, as shown in column (2) of this table, the

Table 3.4: Regression results on non-fully cooperative choices

	First rounds		All rounds		
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Choice_{it}	Choice_{it}	Choice_{it}	Choice_{it}	Choice_{it}
Comp	0.007	0.175	0.999	−2.830***	−2.968***
	(0.546)	(0.417)	(0.911)	(0.269)	(0.462)
Choice_{jt-1}				0.583***	0.582***
				(0.031)	(0.028)
Comp*Choice_{jt-1}				0.202***	0.192***
				(0.033)	(0.031)
Match		0.109***			0.017
		(0.018)			(0.033)
Comp*Match		−0.019			0.029
		(0.032)			(0.038)
Constant	17.016***	15.909***	16.334***	6.607***	6.453***
	(0.404)	(0.361)	(0.773)	(0.171)	(0.397)
Observations	2,823	2,823	25,061	22,238	22,238
R-squared	0.001	0.021	0.010	0.444	0.446

Notes: This table reports results from linear regressions with standard errors (in parentheses) clustered at the session level. *** (**) [*] indicate that the estimated coefficient is significant at the 1% (5%) [10%] level. Specifications (1) and (2) are based on observations from the first rounds of matches only and specifications (3), (4) and (5) are based on all observations.

average choice significantly increases over the matches.[27]

Next, we consider average non-fully-cooperative choices across all rounds. The evoloution of these choices across matches is shown in the right-hand panel of Figure 3.5.[28] Here, a different behavior emerges. When averages are taken across all instead of just the first rounds of a match, the average non-fully-cooperative choice is higher in Comp than in Subs, $p = 0.301$, see column (3) in Table 3.4).

Next we analyze the adjustments across rounds. During a match, subjects observe the past choice(s) of the matched subject and are likely to adjust their own behavior.

[27] We test whether average choice of subjects who do not play fully cooperatively is the same in the two treatment by using a two-sided Mann-Whitney-U Test. The p-value of the null hypothesis that the average non-fully cooperative choice is the same in the two treatments is 0.25, for both the entire experiment and the second half of the experiment. So we fail to reject the null hypothesis.

[28] The right-hand panel of Figure 3.5 is the same as the right-hand panel of Figure 3.3.

If at least some of the subjects (noisily) best-respond it should be the case that in Comp the estimated response function has a higher slope than in Subs (see Table 4.1). Columns (4) and (5) of Table 3.4 report estimates of the observed response functions. The reported results come from linear regressions where the choice of a player is regressed on the choice of the matched player in the previous round (in the same match) as well as the interaction of the other subject's past choice and a treatment dummy. In column (5) additional controls are included for the match and the interaction between match and treatment. Both columns show that in both treatments subjects (partially) follow each other, and the effect is statistically significant.[29] Importantly, the extent to which subjects follow each other is significantly greater in Comp than in Subs. To illustrate, an increase in the choice by a subject, increases the choice of the matched subject in the next round by 0.58 in Subs and by 0.78 in Comp. The effects are very similar when we control for the match and the interaction between match and treatment.

The positive effect of Comp shown in column (3) of Table 3.4 in combination with the result that the extent to which subjects follow each other is greater in Comp than in Subs (cf. columns (4) and (5) in Table 3.4), suggest that at least some subjects try to induce cooperation, to which others (noisily) best-respond. If a subject who increases one's choice above the static Nash equilibrium, for example, with the intention to move towards full cooperation, is matched with a (noisily) best-responding subject or a spiteful subject, choices in the pair will on average end up to be higher (more cooperative) in Comp than in Subs, which is exactly what we observe. This is the mechanism behind our heterogeneity conjecture.

Summarizing, when we focus on non-fully-cooperative choices, we find that behavior is in agreement with the mechanism underlying the heterogeneity conjecture, so that the average (non-fully-cooperative) choice tends to be higher in Comp than in Subs.

We end this section by presenting the results of regressions of treatment effects and responses of subjects using all choices, so including those in the fully-cooperative range. Table 3.5 summarizes the results. The first specification (in column (1)) shows the aggregate (non-significant) treatment effect on choices. The specifications in columns (3) and (4) show the estimated response of subjects to the matched subject's choice, as well as the treatment effect on this response (with and without controlling for the match). As can be seen, the estimated responses are qualitatively similar to those shown in Table 3.4. The size of the estimated response is larger now, because fully cooperative choices as well as subjects responding to full cooperation by fully cooperating themselves are included as well.

[29]Reaction functions being positively sloped in both treatments can be explained by endogenous complementarity that arises when subjects use reciprocal strategies (see also PS).

Table 3.5: Regression results on choice

	(1)	(2)	(3)	(4)
VARIABLES	Choice_{it}	Choice_{it}	Choice_{it}	Choice_{it}
Comp	−0.386	−0.365	−2.421***	−2.190***
	(0.557)	(0.853)	(0.179)	(0.200)
Choice_{jt-1}			0.743***	0.734***
			(0.012)	(0.010)
Comp*Choice_{jt-1}			0.129***	0.126***
			(0.013)	(0.013)
Match	0.208***			0.0602***
	0.041			(0.015)
Comp*Match	0.001			−0.015
	0.060			(0.018)
Constant	16.951***	19.210***	5.005***	4.530***
	(0.487)	(0.703)	(0.118)	(0.096)
Observations	33,024	33,024	29,824	29,824
R-squared	0.043	0.001	0.604	0.607

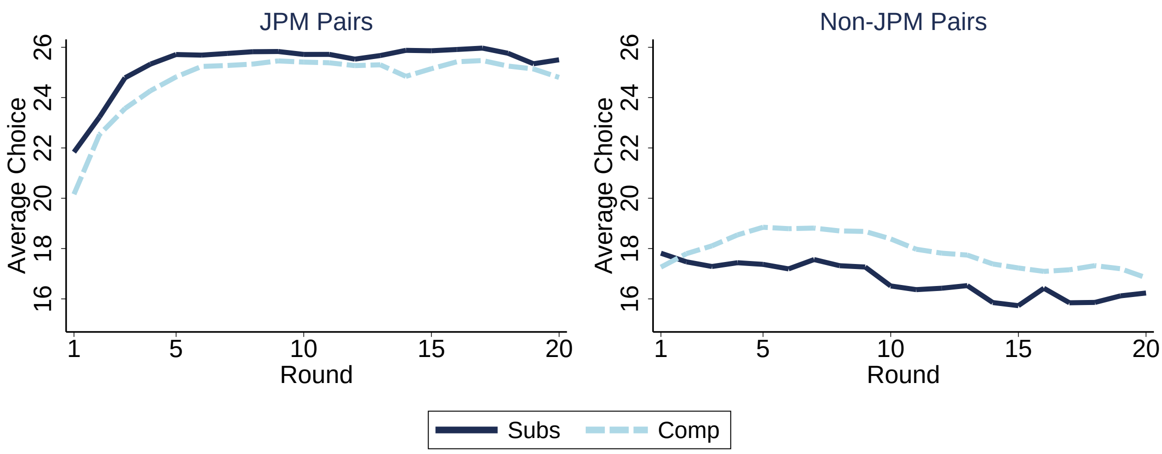
Notes: This table reports results from linear regressions with standard errors (in parentheses) clustered at the session level. *** (**) [*] indicate that the estimated coefficient is significant at the 1% (5%) [10%] level. The dependent variable is a subject's choice in all specifications.

3.4.3. Cooperative versus Non-Cooperative Pairs

In this section we look at the experimental data from a different angle by focusing on the evolution of choices and cooperation at the pair level *within* matches. To do so, we slice up the pairs into those in which the two players succeed in maximizing joint payoff and those in which the two players do not succeed in doing so (along the lines of PS). We classify a pair to be collusive if both subjects choose a number in the interval $[25, 26]$ in at least 60% of the rounds in their individual match. This threshold may look rather low, but if we do not choose the threshold sufficiently low, given that many pairs only play few rounds, they would easily be classified as non-JPM pairs.[30] For example, in order to classify pairs that only play 3 rounds in total in the indefinitely repeated game as JPM pairs if they maximize joint payoff in 2 out of 3 rounds, we need to put the threshold below 66.66%. In any case, any of the qualitative conclusions that are made

[30]Figure 3.15 in the Appendix illustrates the distribution of realized match lengths in the experiment.

Figure 3.6: JPM vs Non-JPM Pairs (60%)



Notes: This figure shows the evolution of choices across matches for JPM and non-JPM pairs respectively on the right- and left-hand sides. A pair is referred to as JPM if both subjects makes a choice in the interval $[25, 26]$ in at least 60% of the rounds in their individual match.

in this section, are robust to changes in this threshold.[31]

Figure 3.6 illustrates the evolution of average choices over time under strategic complements and strategic substitutes for JPM and non-JPM pairs respectively on the left- and right-hand panels. This graph suggests that different choice patterns emerge between JPM and non-JPM pairs. The left-hand panel of Figure 3.6 shows that in Subs pairs who succeed in full cooperation in at least 60% of a match, play higher choices than those in Comp. In the first rounds, the average choice of JPM pairs in Comp is 20, while it is 22 in Subs. As subjects gain experience over time the difference between the treatments disappears. That is, in both treatments once subjects reach the fully cooperative level they remain there. After round 5, the average choice in both treatments is around 25.[32]

The right-hand panel of Figure 3.6 illustrates the evolution of average choice of non-JPM pairs over time. Here we observe that the average choice is higher in Comp than in Subs, $p = 0.303$, see column (1) in Table 3.8). In both treatments the average choice follows a decreasing trend over time. The estimated effect of round on the average choice is -0.061 with $p = 0.004$, see column (2) in Table 3.8). In the Appendix we present figures similar to Figure 3.6 for different thresholds of mutual cooperation, namely from 65% to 80%, where we observe that the difference between the treatments stay the same as the one discussed in this section.

[31] In the Appendix, we include a number of figures in which this threshold is varied (see Figures 3.11 to 3.14).

[32] Table 11 summarizes the average choice for JPM and non-JPM pairs in the first and all rounds of the first match, all matches and the last 10 matches.

3.4.4. Learning across games

In this section we explore the effect of learning across matches. To do so, we study how the behavior of a subject in the first round of a match is affected by (a) the behavior of the partner in the previous match, (b) the length of the previous match, and (c) a subject's own behavior in the previous match (in the spirit of Dal Bó and Fréchette (2011)). We do this in two ways. We check how the variables just mentioned affect a typical subject's probability to start a match fully cooperatively (i.e., by making a choice in the interval [25,26], and how these variables affect a subject's level of choice. We present the results in Table 3.6).

In column (1) and (2) of Table 3.6 we report, for each treatment separately, the results from probit regressions where the dependent variable is a dummy which equals 1 if the choice in the first round of a match is fully cooperative and 0 otherwise. In column (3) and (4) regression results are reported where the dependent variable is a subject's choice in the first round of a match. In all specifications we use the same independent variables: a dummy indicating whether or not the partner in the previous match made a fully cooperative choice in the first round of the previous match, the length of the previous match, and a dummy indicating whether a subject himself made a fully cooperative choice in the very first round of the experiment.

Consider first columns (1) and (2) of Table 3.6. A subject who was matched with someone who played fully cooperatively in the first round of the previous match is more likely to start the current match fully cooperatively in both treatments. However, this effect is statistically significant only in treatment Comp in which cooperation is more risky. Furthermore, in both treatments there is a positive and significant relationship between the length of the previous match and subjects' likelihood of starting a match fully cooperatively. This suggests that after a longer match, during which mutual cooperation is more likely to develop, subjects more often take the risk to start the new match fully cooperatively than after a shorter match. Lastly, a subject who fully cooperated in the first round of the first match in the experiment is more likely to start a new match by full cooperation than someone who did not start the experiment by full cooperation. However, this result is much more pronounced and statistically significant in treatment Subs.

Consider next columns (3) and (4) in Table 3.6 where we report the estimates from linear regressions in which the dependent variable is a subject's choice in the first round of a match. We find that subjects start a match with a higher choice if the partner in the previous match fully cooperated in the first round of the previous match, with the effect being more pronounced in treatment Comp. Also, subjects make higher or more

Table 3.6: Learning across matches

	FullCoop_{it}		Choice_{it}	
	(1)	(2)	(3)	(4)
	Comp	Subs	Comp	Subs
Partner cooperated in	0.065***	0.038	1.609*	0.633
round 1 of previous match	(0.016)	(0.036)	(0.662)	(0.328)
Previous match length	0.001***	0.001**	0.013	0.044**
	(0.001)	(0.001)	(0.008)	(0.014)
Subject cooperated in	0.053	0.271***	−0.293	4.972***
round 1 of match 1	(0.059)	(0.098)	(0.465)	(0.911)
Constant			17.319***	17.673***
			(0.542)	(0.441)
Observations	1,520	1,520	1,520	1,520

Notes: Column (1) and (2) report marginal effects from probit regressions with delta-method standard errors (in parentheses) clustered at the session level. The dependent variable is a dummy which is equal to 1 if the choice is fully cooperatively and 0 otherwise. Column (3) and (4) report results from linear regression with standard errors (in parentheses) clustered at the session level. The dependent variable is a subject's choice. *** (**) [*] indicate that the estimated coefficient is significant at the 1% (5%) [10%] level.

cooperative choices after a longer previous match, where this effect is significant only in treatment Subs. Finally, a subject in treatment Subs makes significantly higher choices if he had chosen a fully cooperative choice himself in the first round of the first match. In treatment Comp, however, the effect is negative and insignificant.

To sum up, a partner's full cooperation in the previous match and a longer previous match increase the choice and the likelihood of full cooperation in the next match. Also, subjects who fully cooperate in the very first round of the experiment are significantly more likely to cooperate later in the experiment in Subs but not so in treatment Comp. Taken together, these results suggest that subjects' behavior is influenced by learning across matches as well as by the nature of the game being being played (complements or substitutes).

3.5. Summary and Discussion

In our experiment subjects play indefinitely repeated dilemma games of strategic substitutes or complements. We find that the full cooperation rate (the percentage of choices at the joint-payoff maximum) is significantly higher under strategic substitutability than under strategic complementarity. We show that this is because under substitutes subjects more often take the risk to initiate full cooperation at the beginning of each repeated game. Moreover, they do so increasingly the more repeated games they play. To illustrate, in the second half of the experiment, the percentage of full cooperation in the first periods has increased to a level above 20%. In contrast, under complements, subjects rarely take this risk, and the percentage remains at about 5% in the second half.

The result that the full cooperation rate is significantly higher under substitutes than under complements goes against the SPNE conjecture, but is in line with the riskiness conjecture. The riskiness conjecture takes into account the riskiness of cooperation, referring to, loosely speaking, how much a player can lose by cooperating in case the other player defects. In the context of this paper, with strategic substitutes it is less risky to fully cooperate or initiate full cooperation than with strategic complements. In this sense, our result is in line with experiments on indefinitely repeated prisoner's dilemma games (Blonski et al., 2011; Dal Bó and Fréchette, 2011) and on coordination games that have shown that payoff-dominant actions are chosen less frequently if they involve more strategic risk (Van Huyck et al., 1990; Schmidt et al., 2003).

Next, if we focus on choices of subjects who do not make choices equal to the joint-payoff maximum, we find that, on average, choices tend to be less cooperative (lower) under strategic substitutes than under complements. This result (although not statistically significant) is in line with the heterogeneity conjecture that posits that under heterogeneity aggregate outcomes tend to be different depending on the strategic environment.

The latter result is in line with behavior observed in finitely repeated games of strategic complements and substitutes of PS who use the same payoff functions as we do here. However, in contrast to PS, we do not find that, in the aggregate, this leads to choices being less cooperative under strategic substitutes than under complements. In our experiment, two opposing behaviors—more frequent play of the joint-payoff maximum and less cooperative "other" choices under substitutes as compared to complements—cancel each other out, so that, on average, choices are not significantly different between the two types of strategic environments.

We suggest that the difference between our results and PS is driven by a difference in a characteristic of the game (known versus unknown end), or the combination of both. The reasoning goes as follows. The repeated game in PS is a long finitely repeated game. It is played with the same partner for 30 rounds, and subjects know this. In contrast, in our experiment, subjects repeatedly play the repeated game with different partners but subjects do not know when it will end. The fundamentals of the interactions are thus very different. In the repeated game of PS, full cooperation, if it occurs, is typically built up gradually: subjects gradually increase their choice towards the level that maximizes joint payoffs. To illustrate, it often takes around 10 rounds to get to this level. In addition, subjects do not initiate full cooperation more frequently in the games with strategic substitutes as compared to those with complements. In our indefinitely repeated games, such gradual build-up is difficult to obtain: subjects do not know how long the repeated game will last, and the expected length is smaller. As compared to PS, full cooperation (if it occurs) hinges more on subjects taking the risk to fully cooperate in the first round (of each repeated game). Therefore, we suspect that the higher relative risk inherent in the games of strategic complements as compared to substitutes has played a fundamental role in our experiment, and not so in PS.

Finally, our findings square well with Embrey et al. (2014). This paper studies in an experiment the effect of strategic commitment on cooperation in indefinitely repeated games of strategic complements and substitutes. Subjects choose an initial action and a strategy (a "machine") at the beginning of each repeated game. Treatments vary with respect to the level of commitment, that is, the costs at which strategies can be adjusted in each round of the repeated game. The treatments vary as well with respect to the strategic environment, with joint payoff maximization being relatively more risky under strategic complements than under strategic substitutes.[33] Interestingly, subjects choose more often joint-payoff maximizing actions under strategic substitutes than under complements when the level of commitment is high, whereas the opposite holds when the level of commitment is low. Relative risk of joint-payoff maximization thus seems to be a dominant force when the level of commitment is high, but not so when it is low.

In summary, our findings and the findings of PS and Embrey et al. (2014) seem to suggest that the effect of the strategic environment on cooperation in repeated games depends on the expected "weight" the relative risk of joint-payoff maximization has on behavior of players. In relatively short games with an unknown end and relatively large differences in this risk, or in games with high levels of commitment, an environment of strategic substitutes seems to be more conducive of cooperation. In long repeated games with a known end, or in games with low commitment and relatively small differences in this risk, less cooperation can be expected under strategic substitutes than under strategic complements.

[33]The difference in minimum thresholds above which full cooperation can be sustained between the two treatments is not as large as in our experiment, though. To illustrate, the minimum thresholds are $\delta^*_{\mathsf{Comp}} = 0.77$ and $\delta^*_{\mathsf{Subs}} = 0.58$ (compared to $\delta^*_{\mathsf{Comp}} = 0.870$ and $\delta^*_{\mathsf{Subs}} = 0.518$, in our experiment).

Appendices

3.A. Instructions

You are participating in an experiment on decision making. You are not allowed to talk or try to communicate with other participants during the experiment. If you have a question, please raise your hand.

Description of the Experiment

In this experiment you will be asked to make a decision in several periods. You will be randomly paired with another participant for a sequence of periods. Each sequence of periods is referred to as a match.

The length of a match is randomly determined. After each round, there is a 90% probability that the match will continue for at least another round. So, for instance, if you are in round 2 of a match, the probability there will be a third round is 90 % and if you are in round 9 of a match, the probability there will be another round is also 90%.

Once a match ends, you will be randomly paired with another participant for a new match.

In each round you and the other participant you will be matched with (referred to as the "other") will be asked to choose a number between 0.0 and 28.0 (in 0.1 steps). The following table gives information about your earnings for some combinations of your and the other's choice. Every participant is given the same table.

You can calculate your and the other's earnings in more detail (for choices that are not multiples of 2 for instance) by using the EARNINGS CALCULATOR on your screen. By filling in a hypothetical value for your own choice and a hypothetical value for the other's choice you can calculate your and the other's earnings for this combination of choices.

Once you have made up your mind, you will enter your decision under DECISION ENTRY and then clicking the button ENTER. In each round you have about 1 minute to enter your decision.

Starting with round 2 of a match, you will be given information about the previous round on your screen. That is, you will be informed about your own and the other participant's choice and your own earnings in points in the previous round.

The identity of the other participants you will be matched with will be unknown to you.

At the end of the experiment you will be paid your earnings in cash and in private. Your total earnings in points are the sum of your earnings in points over all periods of all matches of the experiment. Your earnings in points will be converted into EUR according to the following rate: 300 points = 1 EUR.

Summary

The experiment will consist of a sequence of matches. Each match will consist of a sequence of periods. The number of periods of each match is determined randomly by the computer. After each round, with probability 90% the match continues to another round. You will interact with the same participant for an entire match. After a match is finished, you will be randomly matched with another participant. In each round of a match, you and the other participant you are matched with will choose a number between 0.0 and 28.0 simultaneously.

Payoff tables

Figure 3.7: Payoff table handed out to subjects in the Comp treatment.

						The	Other's	Choice	→							
		0.0	2.0	4.0	6.0	8.0	10.0	12.0	14.0	16.0	18.0	20.0	22.0	24.0	26.0	28.0
	0.0	-28.00	-27.96	-27.87	-27.74	27.57	-27.35	-27.09	-26.78	-26.43	-26.04	-25.60	-25.12	-24.59	-24.02	-23.41
	2.0	-18.16	-17.46	-16.72	-15.93	-15.09	-14.21	-13.29	-12.33	-11.32	-10.26	-9.16	-8.02	-6.84	-5.61	-4.33
	4.0	-10.55	-9.19	-7.78	-6.33	-4.84	-3.30	-1.72	-0.09	1.58	3.29	5.05	6.85	8.70	10.59	12.52
	6.0	-5.16	-3.14	-1.08	1.03	3.19	5.39	7.63	9.91	12.24	14.62	17.04	19.50	22.00	24.55	27.15
	8,0	-2.00	0.68	3.41	6.18	8.99	11.85	14.75	17.70	20.69	23.72	26.80	29.92	33.09	36.30	39.55
	10.0	-1.06	2.28	5.67	9.10	12.57	16.09	19.65	23.26	26.91	30.60	34.34	38.12	41.95	45.82	49.73
Your	12.0	-2.34	1.66	5.70	9.79	13.93	18.11	22.33	26.59	30.90	35.26	39.66	44.10	48.58	53.11	57.69
Choice	14.0	-5.85	-1.19	3.52	8.27	13.06	17.90	22.78	27.71	32.68	37.69	42.75	47.85	53.00	58.19	63.42
↓	16.0	-11.58	-6.26	-0.90	4.51	9.97	15.47	21.01	26.59	32.22	37.90	43.62	49.38	55.18	61.03	66.93
	18.0	-19.54	-13.56	-7.53	-1.46	4.65	10.81	17.01	23.26	29.55	35.88	42.26	48.68	55.15	61.66	68.21
	20.0	-29.72	-23.08	-16.39	-9.66	-2.89	3.93	10.79	17.70	24.65	31.64	38.68	45.76	52.89	60.06	67.27
	22.0	-42.12	-34.82	-27.48	-20.09	-12.65	-5.17	2.35	9.91	17.52	25.18	32.88	40.62	48.40	56.23	64.11
	24.0	-56.75	-48.79	-40.78	-32.73	-24.64	-16.50	-8.32	-0.09	8.18	16.49	24.85	33.25	41.70	50.19	58.72
	26.0	-73.60	-64.98	-56.32	-47.61	-38.85	-30.05	-21.21	-12.33	-3.40	5.58	14.60	23.66	32.76	41.91	51.11
	28.0	-92.68	-83.40	-74.07	-64.70	-55.29	-45.83	-36.33	-26.78	-17.90	-7.56	2.12	11.84	21.61	31.42	41.27

Figure 3.8: Payoff table handed out to subjects in the Subs treatment.

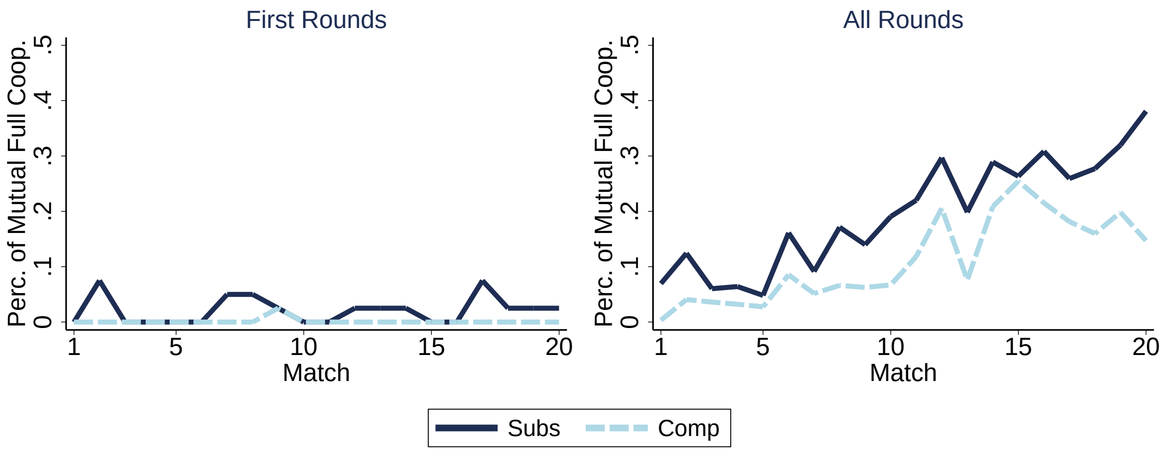
		The Other's Choice →														
		0.0	2.0	4.0	6.0	8.0	10.0	12.0	14.0	16.0	18.0	20.0	22.0	24.0	26.0	28.0
	0.0	-28.00	-22.88	-17.57	-12.09	-6.42	-0.57	5.47	11.68	18.08	24.66	31.42	38.37	45.49	52.80	60.29
	2.0	-22.39	-17.46	-12.35	-7.06	-1.58	4.07	9.91	15.93	22.14	28.52	35.09	41.84	48.77	55.89	63.19
	4.0	-17.43	-12.70	-7.78	-2.69	2.59	8.06	13.70	19.53	25.54	31.73	38.11	44.66	51.40	58.32	65.43
	6.0	-13.13	-8.59	-3.87	1.03	6.12	11.39	16.84	22.47	28.29	34.29	40.47	46.83	53.37	60.10	67.01
	8,0	-9.48	-5.14	-0.61	4.10	8.99	14.07	19.32	24.76	30.38	36.19	42.17	48.34	54.69	61.23	67.94
	10.0	-6.49	-2.34	2.00	6.51	11.21	16.09	21.15	26.40	31.83	37.43	43.23	49.20	55.36	61.70	68.22
Your	12.0	-4.15	-0.19	3.95	8.27	12.77	17.46	22.33	27.38	32.61	38.03	43.63	49.41	55.37	61.51	67.84
Choice	14.0	-2.46	1.30	5.24	9.37	13.68	18.17	22.85	27.71	32.75	37.97	43.37	48.96	54.72	60.67	66.81
↓	16.0	-1.43	2.14	5.89	9.82	13.94	18.24	22.72	27.38	32.22	37.25	42.46	47.85	53.43	59.18	65.12
	18.0	-1.06	2.32	5.88	9.62	13.54	17.64	21.93	26.40	31.05	35.88	40.90	46.10	51.48	57.04	62.78
	20.0	-1.33	1.85	5.21	8.76	12.49	16.40	20.49	24.76	29.22	33.86	38.68	43.68	48.57	54.24	59.79
	22.0	-2.26	0.72	3.89	7.25	10.78	14.49	18.39	22.47	26.74	31.18	35.81	40.62	45.61	50.78	56.14
	24.0	-3.85	-1.05	1.92	5.08	8.42	11.94	15.64	19.53	23.60	27.85	32.28	36.90	41.70	46.68	51.84
	26.0	-6.09	-3.49	-0.71	2.26	5.40	8.73	12.24	15.93	19.81	23.86	28.10	32.52	37.13	41.91	46.88
	28.0	-8.98	-6.57	-3.99	-1.22	1.73	4.87	8.18	11.68	15.36	19.22	23.27	27.50	31.91	36.50	41.27

3.B. Mutual Fully Cooperative Behavior

It is interesting also to look at mutual fully cooperative behavior. Figure 3.9 depicts the evolution of mutual full cooperation across treatments for the first round and all rounds of matches respectively. In all rounds subjects succeed in mutual play of fully cooperative choices significantly more in Subs treatment compared to Comp treatment. In the first rounds, however, mutual full cooperation rate is very low: almost 0% under Comp and below 10% under Subs. Almost none of the pairs succeed in full cooperation in the first rounds with Comp, while they learn to coordinate in mutual full cooperation within a match.

Table 3.7 presents results from probit regressions on mutual full cooperation rate. The probability that a pair reaches full cooperation mutually is significantly higher in the Subs treatment. In both treatments a pair is significantly more likely to coordinate in mutual full cooperation if they played mutual full cooperation in the previous round. Pairs are significantly more likely to achieve mutual full cooperation when they gain experience both within a match and throughout the experiment. The effect of experience on the likelihood of mutual full cooperation is higher, albeit insignificant, in the Comp

Figure 3.9: Mutual Full Cooperation Rate



Notes: This figure shows the evolution of mutual full-cooperation rate across matches, on the left-hand panel for the first rounds only and on the right-hand panel for all rounds.

Table 3.7: Regression results on mutual full cooperation rates

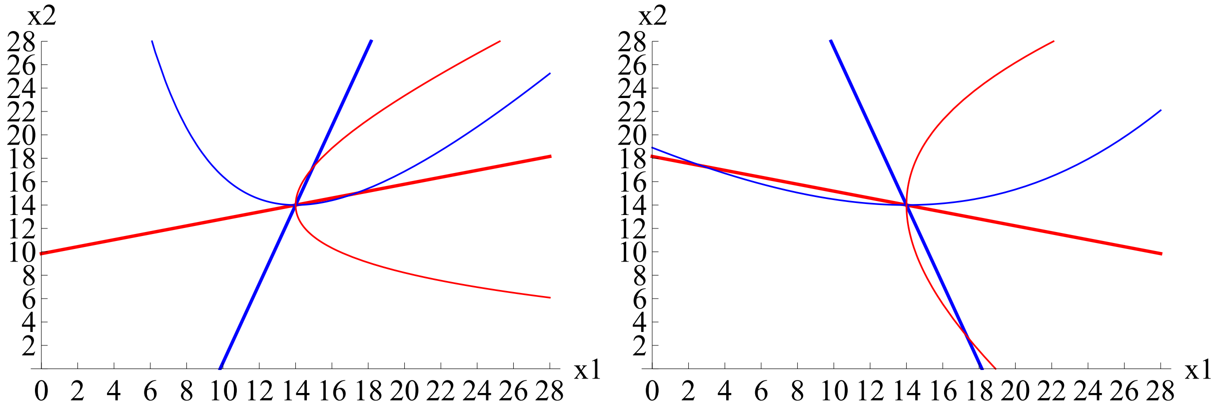
	First rounds		All rounds	
	(1)	(2)	(3)	(4)
VARIABLES	MutualCoop_{it}	MutualCoop_{it}	MutualCoop_{it}	MutualCoop_{it}
Comp	−0.027**	−0.020**	−0.093**	−0.106***
	(0.012)	(0.010)	(0.039)	(0.033)
Match		0.001		0.013***
		(0.001)		(0.002)
Comp*Match		−0.001		0.001
		(0.001)		(0.003)
Observations	3,200	3,200	33,024	33,024

Notes: This table reports marginal effects (delta-method standard errors in parentheses) from probit regression with standard errors clustered at the level of session. The dependent variable is a dummy which is equal to 1 if the choice is mutually fully cooperative and 0 otherwise. *** (**) [*] indicate that the estimated coefficient is significant at the 1% (5%) [10%] level. Specifications (1) and (2) are based on observations from the first rounds only and specifications (3) and (4) are based on all observations.

treatment.

3.C. Additional Graphs and Tables

Figure 3.10: Equilibrium Range



(a) Complements (b) Substitutes

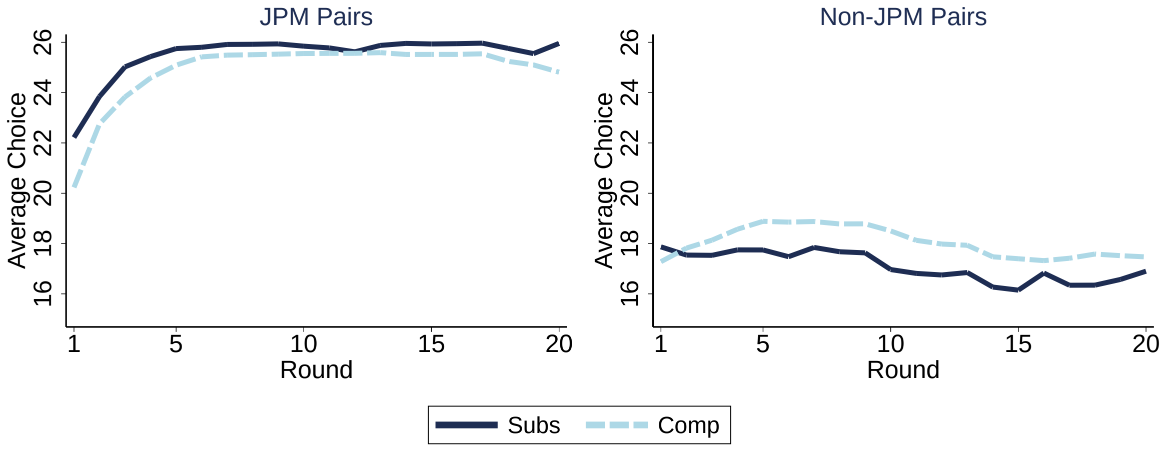
Notes: This figure shows the iso-payoff contours for the games of strategic complements and substitutes respectively on the right- and left-hand panels.

Table 3.8: Regression results on average choice of non-JPM pairs (60%)

	(1)	(2)	(3)
VARIABLES	Choice$_{ij}$	Choice$_{ij}$	Choice$_{ij}$
Comp	0.972	0.797	0.183
	(0.890)	(0.853)	(0.665)
Round		−0.061***	−0.051**
		(0.015)	(0.018)
Comp*Round		0.020	0.027
		(0.028)	(0.027)
Match			0.098
			(0.071)
Comp*Match			0.051
			(0.095)
Constant	17.04***	17.58***	16.50***
	(0.754)	(0639)	(0.398)
Observations	26,500	26,500	26,500
R-squared	0.008	0.013	0.031

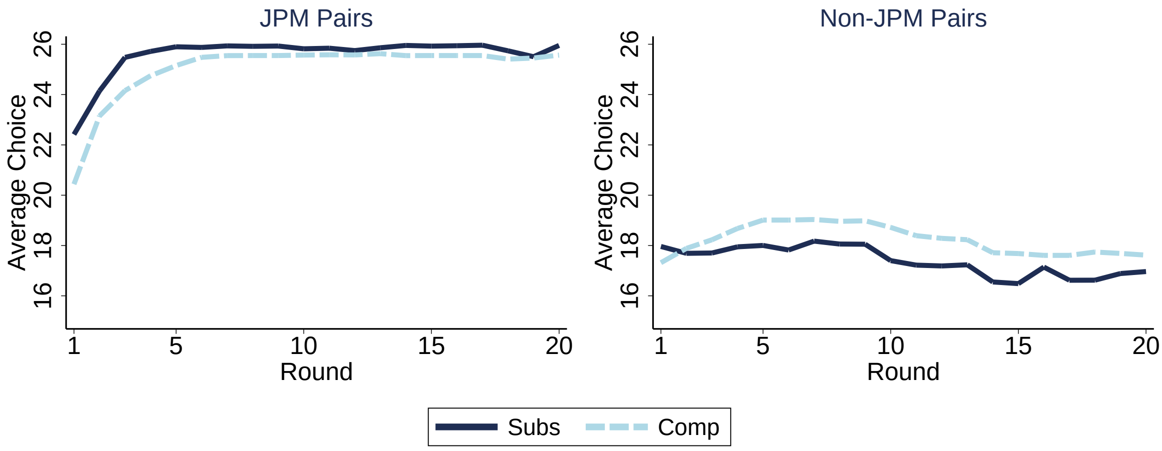
Notes: This table reports results from linear regression with standard errors (in parentheses) clustered at the session level. *** (**) [*] indicate that the estimated coefficient is significant at the 1% (5%) [10%] level.

Figure 3.11: JPM vs Non-JPM Pairs (65%)



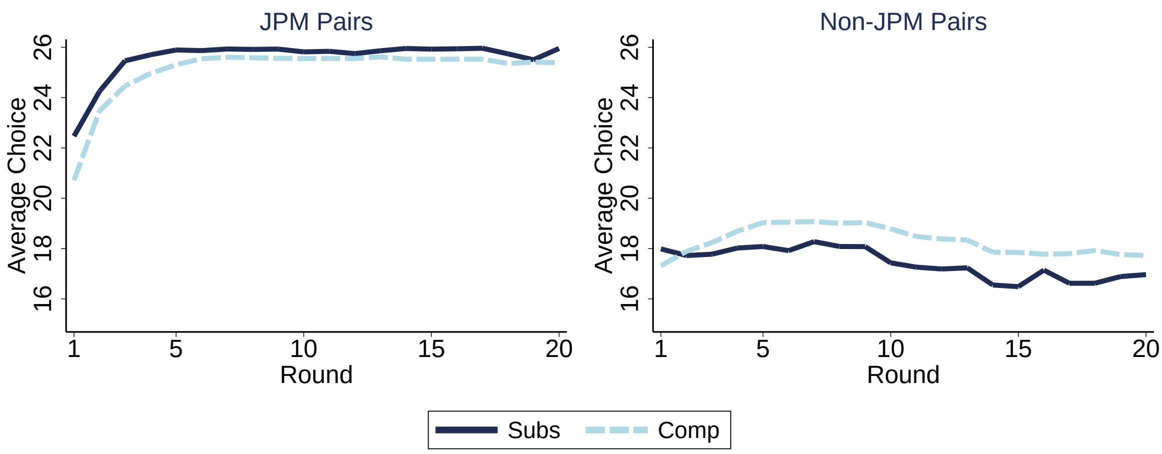
Notes: This figure shows the evolution of choices across matches for JPM and non-JPM pairs. A pair is referred to as JPM if both subjects make a choice in the interval $[25, 26]$ in at least 65% of the rounds in their individual match.

Figure 3.12: JPM vs Non-JPM Pairs (70%)



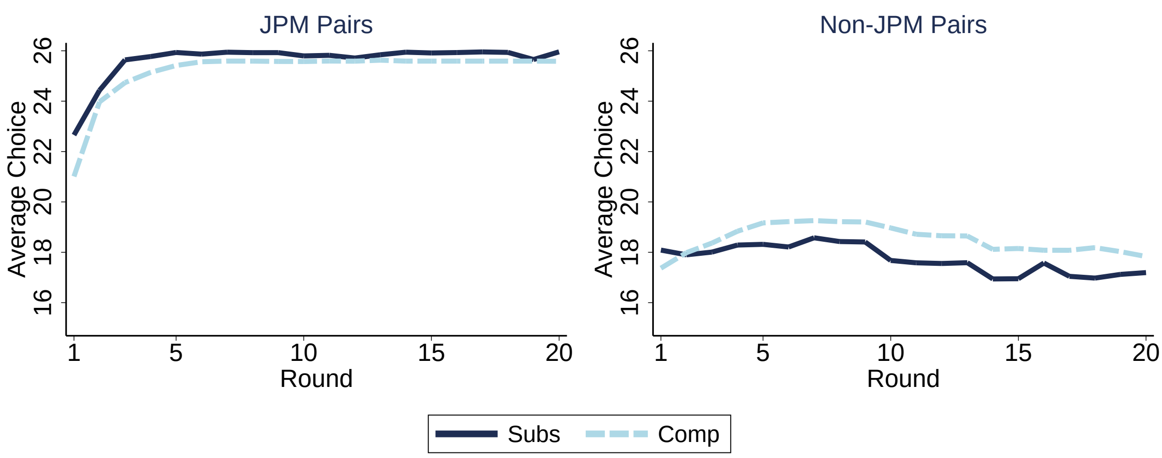
Notes: This figure shows the evolution of choices across matches for JPM and non-JPM pairs. A pair is referred to as JPM if both subjects make a choice in the interval $[25, 26]$ in at least 70% of the rounds in their individual match.

Figure 3.13: JPM vs Non-JPM Pairs (75%)



Notes: This figure shows the evolution of choices across matches for JPM and non-JPM pairs. A pair is referred to as JPM if both subjects make a choice in the interval $[25, 26]$ in at least 75% of the rounds in their individual match.

Figure 3.14: JPM vs Non-JPM Pairs (80%)



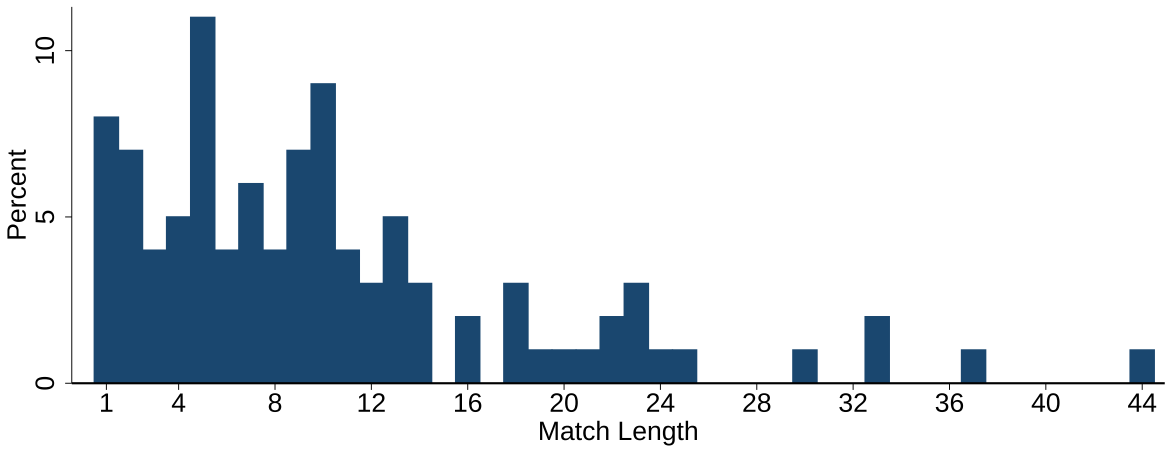
Notes: This figure shows the evolution of choices across matches for JPM and non-JPM pairs. A pair is referred to as JPM if both subjects make a choice in the interval $[25, 26]$ in at least 80% of the rounds in their individual match.

Table 3.9: Regression results on payoffs

VARIABLES	(1) Payoff$_i$	(2) Payoff$_i$	(3) Payoff$_i$
Comp	0.208	−0.246	−1.306
	(1.640)	(1.424)	(1.256)
Round		−0.085	−0.059
		(0.051)	(0.049)
Comp*Round		0.050	0.059
		(0.062)	(0.060)
Match			0.255**
			(0.086)
Comp*Match			0.091
			(0.116)
Constant	33.68***	34.45***	31.45***
	(1.370)	(1.034)	(1.041)
Observations	33,024	33,024	33,024
R-squared	0.001	0.002	0.022

Notes: This table reports results from linear regression with standard errors (in parentheses) clustered at the session level. *** (**) [*] indicate that the estimated coefficient is significant at the 1% (5%) [10%] level.

Figure 3.15: Distribution of Match Lengths in the Experiment



Notes: This figure shows the distribution of the randomly determined match lengths.

Table 3.10: Summary statistics at the individual level

	Average choices					
	First match		All matches		Last 10 matches	
	Comp	Subs	Comp	Subs	Comp	Subs
First round	17.33	17.55	17.50	18.63	17.90	19.37
All rounds	16.31	17.47	18.70	19.09	19.87	20.12

	Full Cooperation Rate					
	First match		All matches		Last 10 matches	
	Comp	Subs	Comp	Subs	Comp	Subs
First round	0.05	0.18	0.05	0.18	0.13	0.25
All rounds	0.06	0.09	0.14	0.27	0.20	0.36

	Average non-fully cooperative choices					
	First match		All matches		Last 10 matches	
	Comp	Subs	Comp	Subs	Comp	Subs
First round	16.08	16.69	17.02	17.03	17.51	17.58
All rounds	16.71	16.55	17.59	16.64	18.33	16.85

Notes: This table summarizes average choices (top panel), full cooperation rates (middle panel) and average non-fully cooperative choices (bottom panel). The results are reported for the first rounds and all rounds of the first match, all and the last 10 matches.

Table 3.11: Summary statistics at the pair level, with 60%

	Average choices in JPM pairs					
	First match		All matches		Last 10 matches	
	Comp	Subs	Comp	Subs	Comp	Subs
First round	-	16	19.63	21.96	19.33	22.30
All rounds	-	23.90	24.64	25.10	24.60	25.18

	Average choices in non-JPM pairs					
	First match		All matches		Last 10 matches	
	Comp	Subs	Comp	Subs	Comp	Subs
First round	17.55	18.26	17.26	17.83	17.55	18.26
All rounds	16.31	16.67	18.16	17.53	18.99	17.92

Notes: This table summarizes average choices for JPM and non-JPM pairs. A pair is referred to as a JPM pair if subjects in this pair sustain full cooperation in at least 60 % of the rounds in that match. We present the averages for the first round and all rounds of the first match, all matches and the last 10 matches separately.

Chapter 4

Choosing To Be Informed in a Repeated Trust Game

4.1. Introduction

Trust has been identified as one of the important determinants of economic growth (Zak and Knack, 2001), the degree of financial development (Guiso et al., 2004), and international trade and investments (Guiso et al., 2008, 2009). This paper contributes to the understanding of the micro-foundations of trust.

Consider a binary trust game where the trustor (he) decides whether to trust or not, and the trustee (she) decides to reciprocate trust or not. In a one-shot trust game, the trustor prefers to be informed about the trustee's type, that is, whether the trustee is trustworthy or not. This information helps him to avoid the potential loss in payoff by trusting an untrustworthy trustee. The situation is different, however, in a repeated trust game. In this case, uncertainty about the trustee's type might help the trustor and trustee to sustain a high level of cooperation and high payoffs. The intuition is that under uncertainty a material-payoff-maximizing ('selfish') trustee has an incentive to be trustworthy today in order to induce trust tomorrow (Kreps et al. (1982)). If this strategy works, the trustee is better off if the trustor is uninformed than if he is informed about her type. Interestingly, also the trustor may be better off being uninformed than informed. Depending on the gains of cooperation, obtaining a 'high' payoff today and a 'low' payoff tomorrow may be better than obtaining a 'medium' payoff twice. If, however, the trustor strongly dislikes getting a lower payoff than the trustee, he will prefer to be informed about the type of the trustee so that he can condition his decision to trust on the latter's type. We design an experiment to study whether trustors *choose* to obtain information about the type of the matched trustee if they have the chance to do so.

In our experiment subjects play a reduced form of a twice-repeated binary trust game under one of two conditions for a number of rounds. The game is a reduced form because it ends at the point the trustor does not trust, or the trustee does not reciprocate trust (Reuben and Suetens, 2014b). In the first condition, referred to as Imperfect, trustors are uninformed about the type of the matched trustee. In the second condition, referred to as Choice, trustors have the choice to obtain information about the type of the trustee. In particular, they have the choice to know whether the trustee will be trustworthy or not on her last move. Because we measure the strategies, we can provide this information to trustors if they want to know. In addition, because we elicit strategies, we can inform all players about the percentage of trustworthy trustees in their population. This is important, because it induces common knowledge about the probability that the trustee is a trustworthy type.

As a theoretical framework we use a model where the trustor is uncertain about the type of the trustee, which can be selfish or trustworthy. The predictions strongly depend on the prior probability that the trustee is a trustworthy type. If this probability is sufficiently high, trustors prefer to be uninformed about the trustee's type, and are predicted to trust today and tomorrow. If the probability is in an 'intermediate' range, they prefer to be informed if they have the choice, which leads them not to trust matched trustees who are selfish types. The same trustors, however, if uninformed, would trust on their first move (today) and with some probability on their second move (tomorrow). Finally, if the probability that the trustee is a trustworthy type is too low, then uninformed trustors do not even trust on their first move.

The predictions also strongly depend on the utility function of trustors, in particular, on the extent to which the trustor (dis)likes an unequal more efficient distribution over an equal less efficient distribution. Following Charness and Rabin (2002), we label this preference 'σ'. A player with a high (positive) σ attaches more weight to efficiency whereas a player with a low (negative) σ dislikes having lower payoffs than the trustee (as in Fehr and Schmidt, 1999). Interestingly, for plausible degrees of inequality aversion ($\sigma < 0$) the trustee will trust if uninformed but prefers to be informed if he has the choice, which destroys trust if the trustee is selfish.

In the experiment we find that in approximately 90% of the cases, subjects in the role of trustor choose to be informed about the matched trustee's type. In line with the above prediction, we find that first-move (today's) trust rates are significantly lower if trustors can choose to be informed than if uninformed, both between and within subjects. For second-move (tomorrow's) trust rates, we do not find a treatment effect. In order to study the relation between trustors' behavior and their σ, we elicited this preference in a dictator game that preceded the trust games. We find that the elicited σ's are negatively correlated with the percentage of times trustors in the experiment choose to be informed. We also find that the within-subjects difference in first-move trust rates between cases where the trustor is uninformed and cases where he is informed that the trustee is selfish is larger, the smaller the trustor's elicited σ.

We are not aware of other papers that study in a trust or related game whether trustors choose to be informed about the type of the trustee *ex ante*, that is, before they decide to trust or not. Some papers study whether trustors avoid potentially harmful knowledge in a trust game *ex post*. Examples are Aimone and Houser (2012) and Aimone and Houser (2013). In these studies trustors in a one-shot binary trust game have the option to replace the strategy of the trustee they were initially matched with by the strategy of a trustee randomly drawn from a population of trustees identical to the one they were playing with. Findings are that trustors typically opt into this, and thus avoid getting informed about the matched trustee's choice (type). In addition, trustors trust less if this avoidance option is not provided.

The remainder of the paper is organized as follows. In section 4.2 we present the experimental games and the theoretical prediction. In section 4.3 we discuss the experimental design, procedures, and research questions. In section 4.4 we present the experimental results. Section 4.5 concludes.

4.2. Experimental Games and Predictions

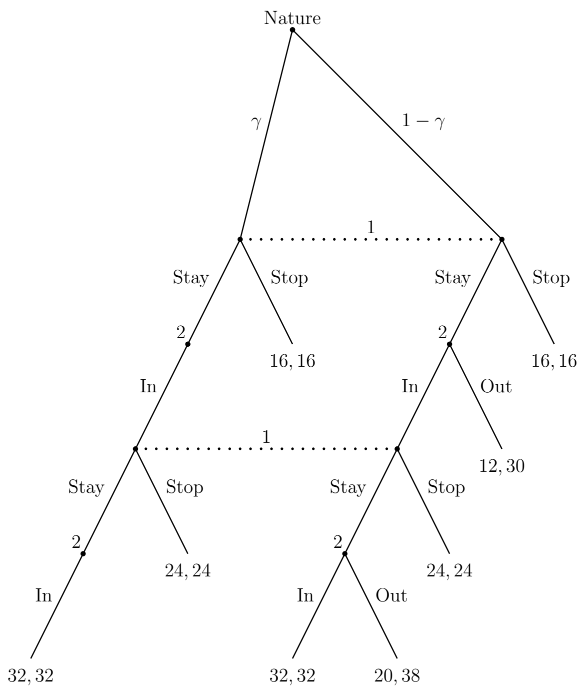
In the experiment subjects play twice-repeated trust games as used by Reuben and Suetens (2014b). In the game the trustor (player 1) decides whether to trust by playing STAY or not to trust by playing STOP. We refer to this choice by player 1's 'first move'. If player 1 plays STOP on his first move, the game ends. If player 1 chooses STAY, the game continues. If the game continues, the trustee (player 2) decides whether to reciprocate by choosing IN or not to reciprocate by choosing OUT. We refer to this choice as player 2's 'first move'. If player 2 chooses OUT, the game ends. If player 2 chooses IN, the game continues. After continuation, player 1 chooses between STAY and STOP, referred to as player 1's 'second move'. If player 1 chooses STOP, the game ends. If player 1 chooses STAY, the game continues to the last period. In the last period, if played, player 2 decides between IN and OUT. This choice is referred to as player 2's 'second move'. In each step, the total surplus increases so that it is maximized if player 1 chooses STAY and player 2 chooses IN.

In the first version of the game, which we refer to as Imperfect, we assume that the trustee can be either a selfish or trustworthy type, and the trustor does not observe the trustee's type. In the second version of the game, which we refer to as Choice, the trustor has the choice to be informed about the type of the trustee. In what follows we describe both experimental games in more detail and we derive the theoretical predictions.

4.2.1. Imperfect Game

Figure 4.1 shows the Imperfect game. The game has 5 periods. In the first period, Nature draws the type of player 2. With probability γ, player 2 is a type who commits to be trustworthy by reciprocating on both moves (so a trustworthy type). With probability $1-\gamma$, player 2 is a selfish type. γ is common knowledge. We use perfect Bayesian equilibrium as a theoretical concept. We follow Camerer and Weigelt (1988) and Anderhub et al. (2002) in deriving the theoretical predictions.[34]

Figure 4.1: Imperfect



From period 2, the players decide in turns to STAY or STOP (player 1), and to stay IN or go OUT (player 2). At the point player 1 chooses STOP, or player 2 chooses OUT, the game ends. To derive the equilibrium predictions, we assume that material payoffs in Figure 4.1 are the true preferences for the players and these preferences are common knowledge.

Player 1 does not observe the type of player 2. Given that γ, which is the probability of player 2 being a trustworthy type, is common knowledge, player 1's prior belief about

[34]For similar games and theoretical predictions, see Neral and Ochs (1992); McKelvey and Palfrey (1992); Brandts and Figueras (2003); Duffy and Munoz-García (2013).

player 2 being a trustworthy type $\mu_1 = \gamma$. If the game continues to period 4, player 1 updates his belief about player 2's type in this period. If this is the case, player 2 must have chosen IN on her first move. This might be because player 2 is a trustworthy type who reciprocates on both moves or because she is a selfish type who only reciprocates on her first move but not on her second. In the following we solve the game.

In the last period, a trustworthy player 2 chooses IN and a selfish player 2 chooses OUT by definition. Player 1 chooses STAY in the last information set if his expected payoff from choosing STAY is larger than his expected payoff from choosing STOP. In the last information set, observing the first move of player 2, player 1 updates his belief about player 2's type using Bayesian updating. Denote the updated belief after player 2 playing IN on her first move by μ_2. Player 1 chooses STAY on his second move if $\mu_2 32 + (1-\mu_2)20 > 24$. That is, player 1 plays STAY on his second move if $\mu_2 > \frac{1}{3}$.

We first consider the case when $\gamma > \frac{1}{3}$. Given that $\mu_1 > \frac{1}{3}$, it also holds that $\mu_2 > \frac{1}{3}$ once player 1 arrives in period 4. In this case, player 1 chooses STAY on his second move. Player 2 plays IN on her first move if she is trustworthy. If player 2 is selfish, then she also chooses IN on her first move since $38 > 30$. Player 1 chooses STAY in the first information set if his expected payoff from choosing STAY is larger than his expected payoff from choosing STOP, that is if $\gamma 32 + (1-\gamma)20 > 16$. This inequality always holds as γ cannot be negative. Thus player 1 chooses STAY on his first move.

We now consider the case when $\gamma < \frac{1}{3}$. We first show that the equilibrium is in mixed strategies in this case. Let m be the probability that player 1 chooses STAY on his second move and p be the probability that a selfish player 2 chooses IN on her first move. Player 1 updates his belief according to Bayes' Rule. His updated belief given that player 2 chooses IN on her first move is given by $\mu_2 = \frac{\gamma}{\gamma + p(1-\gamma)}$. We first show that p cannot take the values 0 and 1 (following Anderhub et al., 2002). Suppose that $p = 0$, then by Bayesian updating, $\mu_2 = \frac{\gamma}{\gamma + 0(1-\gamma)} = 1$. This implies that player 1 optimally plays STAY ($(1-m) = 0$), in which case player 2 would optimally choose IN with probability $p = 1$ contradicting the initial supposition. Thus $p > 0$ should be satisfied in equilibrium. Suppose that $p = 1$. Then by Bayesian updating, $\mu_2 = \frac{\gamma}{\gamma + 1(1-\gamma)} = \gamma < \frac{1}{3}$. This implies that $m = 0$. But then player 2 would choose OUT on her first move, which contradicts the initial supposition that $p = 1$. Thus $0 < p < 1$ should hold. That is, player 2 uses a mixed strategy on her first move. In this case, player 1 should also use a mixed strategy on his second move as otherwise player 2 would optimally play a pure strategy on her first move as well. Therefore, we have proven that $0 < p < 1$ and $0 < m < 1$ when $\gamma < \frac{1}{3}$.

Next we calculate the equilibrium mixing probabilities. Player 1, on his second move, plays STAY with probability m and plays STOP with probability $(1-m)$. He chooses

m such that a selfish player 2 is indifferent on her first move. Thus we have $m38 + (1-m)24 = 30$, yielding $m = \frac{3}{7}$. Player 2, on her first move, plays IN with probability p and plays OUT with probability $(1-p)$. He chooses p such that player 1 is indifferent on his second move, that is $\mu_2 = \frac{\gamma}{\gamma+p(1-\gamma)} = \frac{1}{3}$, yielding $p = \frac{2\gamma}{1-\gamma}$.

Finally, player 1, on his first move, plays STAY if his expected payoff from doing so is larger than his expected payoff from choosing STOP. That is, if $\gamma[m32+(1-m)24] + (1-\gamma)[p(m20+(1-m)24)+(1-p)12] > 16$. Substituting $m = \frac{3}{7}$ and $p = \frac{2\gamma}{1-\gamma}$, we get that this condition holds if $\gamma > \frac{1}{9}$. Thus player 1 chooses STAY if $\frac{1}{9} < \gamma < \frac{1}{3}$ and chooses STOP if $\gamma < \frac{1}{9}$. The equilibrium predictions for the Imperfect game are summarized as follows:[35]

Prediction Imperfect. *Let player 2 be a trustworthy type with probability γ and a selfish type with probability $1-\gamma$. The theoretical predictions of the* Imperfect *game are:*

(a) $\gamma > \frac{1}{3}$: Player 1 plays STAY on his both moves. Beliefs are $\mu_1 = \gamma$ and $\mu_2 = \gamma$. A trustworthy player 2 plays IN on both moves and a selfish player 2 plays IN on her first move and OUT on her second move.

(b) $\frac{1}{9} < \gamma < \frac{1}{3}$: Player 1 plays STAY on his first move and plays STAY with probability $\frac{3}{7}$ and STOP with probability $\frac{4}{7}$ on his second move. Beliefs are $\mu_1 = \gamma$ and $\mu_2 = \frac{1}{3}$. A trustworthy player 2 plays IN on both moves. A selfish player 2 plays IN with probability $\frac{2\gamma}{1-\gamma}$ and plays OUT with probability $1 - \frac{2\gamma}{1-\gamma}$ on her first move and plays OUT on her second move.

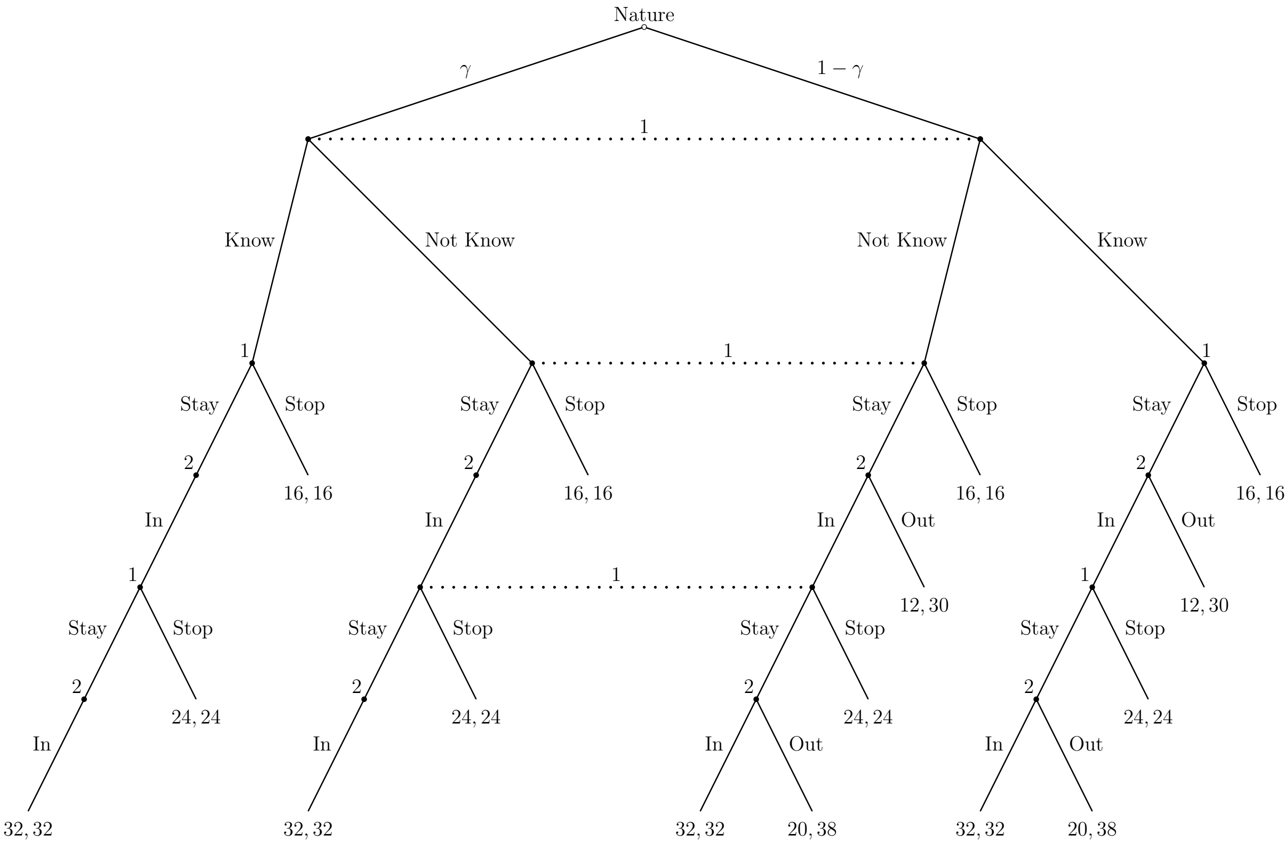
(c) $\gamma < \frac{1}{9}$: Player 1 plays STOP on his first move and plays STAY with probability $\frac{3}{7}$ and STOP with probability $\frac{4}{7}$ on his second move. Beliefs are $\mu_1 = \gamma$ and $\mu_2 = \frac{1}{3}$. A trustworthy player 2 plays IN on both moves. A selfish player 2 plays IN with probability $\frac{2\gamma}{1-\gamma}$ and OUT with probability $1 - \frac{2\gamma}{1-\gamma}$ on her first move and OUT on her second move.

4.2.2. Choice Game

Figure 4.2 shows a game where before player 1 makes his first move in a trust game, he can choose to be perfectly informed about the type of player 2, that is, to know or not to know whether player 2 is a trustworthy type. We refer to this game as the Choice game. It is common knowledge whether player 1 knows or does not know the type of player 2. Note that we continue to refer to the first (second) move of player 1 in the trust game, as player 1's first (second) move.

[35]We do not include cases where $\gamma = \frac{1}{3}$ or $\gamma = \frac{1}{9}$ because in these cases, players are indifferent.

Figure 4.2: Choice



The game has 6 periods. In the first period Nature draws the type of player 2 (trustworthy or selfish). The probability of player 2 being a trustworthy type, γ, is common knowledge. In the second period, player 1 is asked whether he wants to obtain information about the type of player 2. If player 1 chooses KNOW in period 2, he is informed about whether player 2 is a trustworthy or a selfish type. If player 2 chooses NOT KNOW, he does not receive any information about the type of player 2. In periods 3 to 6 the game proceeds in the same way as the above-described trust game. Namely, both players choose in turn whether to STAY or STOP (player 1), or whether to stay IN or go OUT (player 2).

Before calculating the optimal choice of player 1, that is, whether player 1 chooses to KNOW or NOT KNOW the type of player 2, we solve the subgames depending on player 1's choice in period 2.

Player 1 chooses KNOW

Suppose that player 1 chooses KNOW in period 2 and solve the game by backward induction. In the last period of the game, a trustworthy player 2 chooses IN and a selfish player 2 chooses OUT. Player 1, on his second move, chooses STAY if player 2 is trustworthy since $32 > 24$ and chooses STOP if player 2 is selfish since $24 > 20$. A trustworthy player 2 chooses IN on his first move by definition. A selfish player 2 chooses OUT on her first move as $30 > 24$. Player 1, on his first move , chooses STAY if he is matched with a trustworthy player 2 since $32 > 16$ and chooses STOP if he is matched with a selfish player 2 since $16 > 12$. So, in summary, if player 1 chooses KNOW in period 2, a trustworthy player 2 will choose IN on both moves and player 1 chooses STAY on both moves in equilibrium. A selfish player 2 chooses OUT on both moves and player 1 chooses STOP on both moves.

Player 1 chooses NOT KNOW

When player 1 chooses NOT KNOW, we are in the Imperfect game. We refer to section 4.2.1 for the calculation of the equilibrium in this subgame.

Choice to KNOW or NOT KNOW

We now calculate the optimal choice of player 1 in period 2. Player 1 chooses KNOW if his expected payoff from playing KNOW is larger than his expected payoff from NOT KNOW. We consider the 3 ranges for γ that are relevant given that player 1 does NOT KNOW (see 4.2.1): $\gamma > \frac{1}{3}$, $\frac{1}{9} < \gamma < \frac{1}{3}$ and $\gamma < \frac{1}{9}$.

First, we consider the case where $\gamma > \frac{1}{3}$. Player 1 chooses KNOW in this case if

$$\gamma 32 + (1-\gamma)16 > \gamma 32 + (1-\gamma)20.$$

This inequality never holds since 16<20. Thus player 1 chooses NOT KNOW whenever $\gamma > \frac{1}{3}$.

Second, we consider the case where $\frac{1}{9} < \gamma < \frac{1}{3}$. Player 1 chooses KNOW in this case if

$$\gamma 32 + (1-\gamma)16 > \gamma[m32 + (1-m)24] + (1-\gamma)[p(m20 + (1-m)24) + (1-p)12].$$

Plugging in the equilibrium values for m and p derived above, this inequality holds only when $\gamma < \frac{1}{5}$. Thus player 1 chooses KNOW whenever $\frac{1}{9} < \gamma < \frac{1}{5}$ and chooses NOT KNOW whenever $\frac{1}{5} < \gamma < \frac{1}{3}$. He is indifferent between KNOW and NOT KNOW when $\gamma = \frac{1}{5}$.

Last, we consider the case where $\gamma < \frac{1}{9}$. Player 1 chooses KNOW in this case if

$$\gamma 32 + (1-\gamma)16 > 16.$$

This inequality always holds, thus player 1 chooses KNOW whenever $\gamma < \frac{1}{9}$.

The theoretical predictions for the Choice game are summarized as follows:

Prediction Choice. *Let player 2 be a trustworthy type with probability γ and a selfish type with probability $1 - \gamma$. The theoretical predictions of the* Choice *game are:*

(a) $\gamma > \frac{1}{3}$: Player 1 plays NOT KNOW. Player 1 plays STAY on both moves if he plays NOT KNOW in period 2. Beliefs are $\mu_1 = \gamma$ and $\mu_2 = \gamma$. When player 1 plays KNOW, then he plays STAY on both moves when player 2 is a trustworthy type and STOP on both moves when player 2 is a selfish type. A trustworthy player 2 plays IN on both moves when player 1 plays KNOW or NOT KNOW. A selfish player 2 plays IN on her first move and OUT on her second move when player 1 plays NOT KNOW. A selfish player 2 plays OUT on both moves when player 1 plays KNOW.

(b) $\frac{1}{9} < \gamma < \frac{1}{3}$: Player 1 plays NOT KNOW if $\gamma > \frac{1}{5}$ and KNOW if $\gamma < \frac{1}{5}$. If $\gamma = \frac{1}{5}$ he is indifferent between playing KNOW and NOT KNOW. Player 1 plays STAY on his first move and plays STAY with probability $\frac{3}{7}$ and STOP with probability $\frac{4}{7}$ on his second move when he plays NOT KNOW. Beliefs are $\mu_1 = \gamma$ and $\mu_2 = \frac{1}{3}$. When player 1 plays KNOW, he plays STAY on both moves if player 2 is a trustworthy type and STOP on both moves if player 2 is a selfish type. A trustworthy player 2 plays IN on both moves when player 1 plays KNOW or NOT KNOW. A selfish player 2 plays IN with probability $\frac{2\gamma}{1-\gamma}$ and plays OUT with probability $1 - \frac{2\gamma}{1-\gamma}$ on her first move and plays OUT on her last move when player 1 plays NOT KNOW. A selfish player 2 plays OUT on both moves when player 1 plays KNOW.

(c) $\gamma < \frac{1}{9}$: Player 1 plays KNOW. Player 1 plays STOP on his first move and plays STAY with probability $\frac{3}{7}$ and STOP with probability $\frac{4}{7}$ on his second move when he plays NOT KNOW. Beliefs are $\mu_1 = \gamma$ and $\mu_2 = \frac{1}{3}$. When player 1 plays KNOW, he plays STAY on both moves if player 2 is a trustworthy type and STOP on both moves if player 2 is a selfish type. A trustworthy player 2 plays IN on both moves when player 1 chooses KNOW or NOT KNOW. A selfish player 2 plays IN with probability $\frac{2\gamma}{1-\gamma}$ and OUT with probability $1 - \frac{2\gamma}{1-\gamma}$ on her first move and plays OUT on her second move when player 1 plays NOT KNOW. A selfish player 2 plays OUT on both moves when player 1 chooses KNOW.

4.2.3. Non-Standard Preferences of Trustors

Several studies have shown that rather than being material payoff maximizers, individuals sometimes have other motivations. In particular, several individuals dislike realizing a lower material payoff than others (see Fehr and Schmidt (1999), Bolton and Ockenfels

(2000)), or prefer efficiency over inefficient distributions (Charness and Rabin (2002)). In this section we discuss the intuition of how the theoretical predictions are affected if the trustor has such motivations.

Intuitively, it should be clear that the more the trustor dislikes receiving a lower material payoff than the trustee, the higher γ (the probability that the trustee is a trustworthy type) should be for the trustor to trust. In addition, in the case that the trustor can choose to obtain information about the trustee's type, it is also intuitive to see that the more the trustor dislikes receiving a lower material payoff than the trustee, the more likely it is that he will want to have information about player 2's type. This will allow him to avoid unequal payoff distributions. To illustrate, assume that player 1 has the following utility function (Charness and Rabin (2002)):

$$U(x_1, x_2) = x_1 + \sigma(x_2 - x_1), \tag{4.1}$$

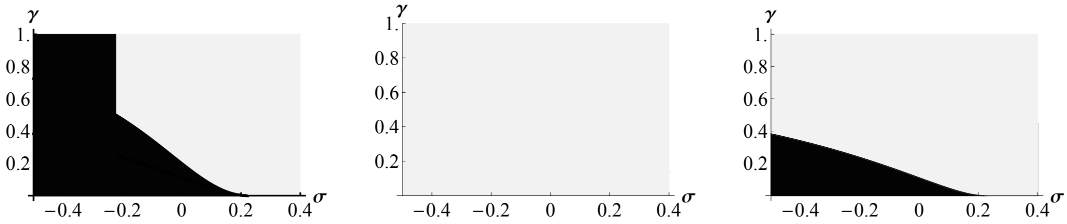
where x_1 and x_2 are the material payoffs for player 1 and player 2.

If $\sigma < 0$, player 1 dislikes being behind player 2 in terms of material payoff. The more negative is σ, the more player 1 dislikes disadvantageous inequality aversion. For some players, however, $\sigma > 0$, because they prefer an unequal distribution with relatively high total payoffs (efficiency) over an equal distribution with relatively low total payoffs (Engelmann and Strobel (2004)). In what follows we discuss how choices of player 1 are affected by using this utility function. For simplicity, we assume that σ is common knowledge. This implies that choices of player 2 follow in a straightforward way from those of player 1. In section A of the Appendix, the full equilibrium predictions and proofs are provided.

Consider first the Imperfect game. If $\sigma < 0$, the threshold for γ above which player 1 plays STAY on his second move gets larger in comparison to the case when $\sigma = 0$. Thus player 1 will less easily play STAY on his second move, and the threshold will depend on his σ. The threshold becomes $\frac{2-9\sigma}{6-9\sigma}$ as compared to $\frac{1}{3} \approx 0.333$. Along the same lines, the threshold for γ above which player 1 plays STAY on his first move also gets larger in comparison to the case when $\sigma = 0$. It becomes $\left(\frac{2-9\sigma}{6-9\sigma}\right)^2$ instead of $\frac{1}{9} \approx 0.111$. To illustrate, a player 1 with $\sigma = -0.2$ ($\sigma = -0.4$) will choose STAY on his first move if $\gamma > 0.237$ ($\gamma > 0.340$) and on his second move if $\gamma > 0.487$ ($\gamma > 0.583$).

If $\sigma > 0$, the threshold for γ above which player 1 plays STAY on his second move gets smaller in comparison to the case when $\sigma = 0$, as well as the threshold above which player 1 plays STAY on his first move. A special case is that $\sigma > \frac{2}{9}$. In this case player 1 cares so much about efficiency that he will unconditionally choose STAY on both moves, independently from γ. The reason is that he then prefers distribution $(20, 38)$

Figure 4.3: Predicted First Moves of Player 1



(a) Choice-Selfish (b) Choice-Trustworthy (c) Imperfect

Notes: The figure illustrates the predicted first moves of player 1. The predictions in Choice when player 2 is selfish are shown in the left-hand panel, in Choice when player 2 is trustworthy are shown in the middle panel and in Imperfect are shown in the right-hand panel. In the black area, player 1 chooses STOP. In the grey area, player 1 mixes between STAY and STOP. In the light grey area, player 1 chooses STAY.

over $(24, 24)$ even when he would know that player 2 is selfish.

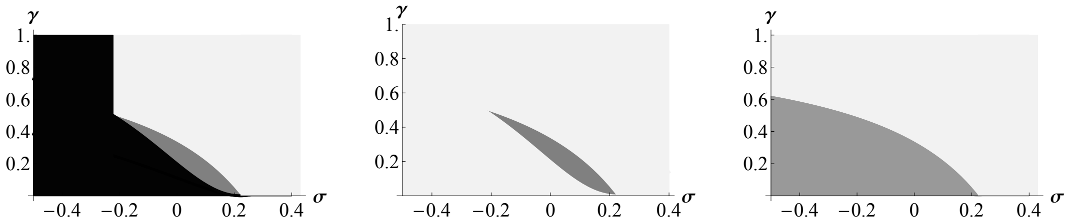
We now turn to the Choice game. Whether player 1 plays NOT KNOW and STAY on his first move depends on σ. Even if the probability the trustee is a trustworthy type is very high, that is, if $\gamma > \frac{2-9\sigma}{6-9\sigma}$, player 1 will prefer KNOW and STOP if he is sufficiently inequality averse. In particular, he prefers KNOW and therefore STOP if $\sigma < -\frac{2}{9} \approx -0.222$. The reason is that in this case he prefers the distribution of payoffs obtained when choosing STOP on his first move $(16, 16)$ over the distribution that would result if he plays STAY on both moves and player 2 is selfish (20,38). If $\sigma > -\frac{2}{9}$, he prefers NOT KNOW and chooses STAY on both moves. In the extreme case that $\sigma > \frac{2}{9}$, player 1 is indifferent between KNOW and NOT KNOW and therefore chooses STAY on both moves whenever $\gamma > \frac{2-9\sigma}{6-9\sigma}$. The reason is that in this case player 1 prefers $(20, 38)$ over $(32, 32)$.

When $\gamma < \frac{2-9\sigma}{6-9\sigma}$, the predictions in the Choice game have the same intuition as those where $\sigma = 0$ (of course, provided that $-\frac{2}{9} < \sigma < \frac{2}{9}$). The difference is that the threshold for γ influencing player 1's choice to KNOW, now also depends on σ. If $\sigma < 0$ ($\sigma > 0$), player 1 will prefer to KNOW more (less) easily than when $\sigma = 0$ provided that $-\frac{2}{9} < \sigma < \frac{2}{9}$.

Figure 4.3 and 4.4 gives an overview of the predicted first and second move choices of player 1 as a function of γ and σ.[36]

[36] These predictions integrate the choice of player 1 on receiving information or not about the type of player 2. If $\sigma < -\frac{2}{9}$ player 1 plays KNOW for any value of γ leading to a black area for this range. If $\sigma > \frac{-2}{9}$ player 1 plays KNOW for $\gamma < \frac{(2-9\sigma)^2}{20-36\sigma+81\sigma^2}$, and plays NOT KNOW for $\gamma > \frac{(2-9\sigma)^2}{20-36\sigma+81\sigma^2}$.

Figure 4.4: Predicted Second Moves of Player 1



(a) Choice-Selfish (b) Choice-Trustworthy (c) Imperfect

Notes: The figure illustrates the predicted second moves of player 1 given that he has chosen STAY on his first move. The predictions in Choice when player 2 is selfish are shown in the left-hand panel, in Choice when player 2 is trustworthy are shown in the middle panel and in Imperfect are shown in the right-hand panel. In the black area, player 1 chooses STOP. In the grey area, player 1 mixes between STAY and STOP. In the light grey area, player 1 chooses STAY.

4.3. Research Methods and Questions

4.3.1. Experimental Design

The experiment was designed in order to implement the above-described games as close as we could. The experiment has two treatments, labeled according to the implemented games. In treatment Imperfect, the game shown in Figure 4.1 was induced, and in treatment Choice, the game shown in Figure 4.2 was induced. Subjects played each of the two games a number of times to allow for learning. After each round they were randomly rematched. Equilibrium theory would be an unreasonable theoretical framework, if there is no scope for learning.

We measured choices in the experiment by eliciting strategies.[37] We asked subjects in the role of player 1 whether they choose STOP on their first move, STAY on their first move and STOP on their second move, or STAY on both moves. Likewise, we asked subjects in the role of player 2 whether they choose OUT on their first move, IN on their first move and OUT on their second move, or IN on both moves. We elicited strategies in order to have information about the type of subjects in the role of player 2. In particular, subjects who choose IN on both moves in the role of player 2, were classified as trustworthy types.[38]

We elicited the strategies of subjects in the role of player 2 in two steps. In a first step, at the beginning of the experiment, we asked the subjects whether they decide to play IN-IN or not IN-IN. Subjects were committed to this choice for the entire experiment,

[37]As shown by Reuben and Suetens (2012, 2014b), trust and reciprocation rates based on 'cold' elicitation are not significantly different from those based on 'hot' elicitation.

[38]Instructions handed out to the subjects are included in section B of the Appendix.

and they were aware of this. The aim was twofold. First, we wanted to measure the percentage of subjects in each session in the role of player 2 that commit to being trustworthy by choosing IN-IN, namely γ. Given that the theoretical predictions strongly depend on γ, this occurred to us as crucial. In order to induce common knowledge of γ, all subjects in the session were informed about the ratio of subjects who chose IN-IN (for example, 6 out of 16 players 2 chose IN-IN). Second, we needed the type of player 2 to be communicable to player 1 in Choice.

In a second step, which was, in contrast to the first step, repeated in each round of the experiment, subjects were asked to submit the remainder of their strategy. Specifically, subjects in the role of player 1 were asked to submit their strategy (which could be STOP, STAY-STOP, or STAY-STAY), and so were subjects in the role of player 2 who had *not* chosen IN-IN (which could be OUT, or IN-OUT).[39]

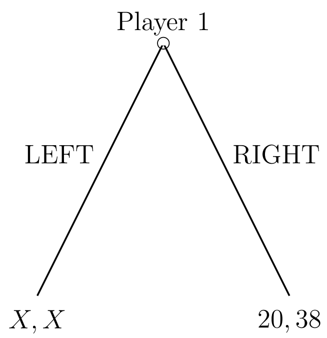
Subjects in both treatments went through these two steps. In Choice, in between these two steps, subjects in the role of player 1 were given the choice to be perfectly informed about the matched player 2. In particular, player 1 was given the choice to know whether the matched player 2 chose IN-IN in step 1 or not (that is, player 2's type). If he chose to click on the button to know, he received the message "The matched player 2 has chosen IN-IN." in the case the matched player 2 was a trustworthy type, or the message "The matched player 2 has not chosen IN-IN." in the case the matched player 2 was a selfish type, next to the share of subjects who chose IN-IN. If he did not want to know, he did not get the additional information. In addition, subjects in the role of player 2 in Choice were informed about whether the matched player 1 was perfectly informed about their type or not. So at the point selfish player 2 types decided whether to choose IN-OUT or OUT, they knew that the matched player 1 knew they would not reciprocate on their last move, if player 1 had chosen to be informed. Also, subjects in the role of player 1 were informed that the matched player 2 would be informed if they would choose to KNOW.

Once all subjects had submitted their decisions, the computer matched the decisions of player 1 and player 2 and calculated the payoffs. At the end of each round, payoffs of both players were displayed on the screen. After each round subjects were randomly rematched.

Obviously, our implementation of a trustworthy type is a simplification. There are a number of reasons why we use this 'short-cut', and did not, for example, give subjects in the role of player 1 the option to acquire information about the past behavior of subjects in the role of player 2. The first reason is that we wanted to induce (perfect information

[39]Subjects in the role of player 2 who had chosen IN-IN in the first step had no decision to make in the remaining of the experiment.

Figure 4.5: Dictator Game



Notes: Figure depict the dictator game played at the beginning of the experiment.

about) player 2's type as typically defined in theory as closely as possible without turning to 'robots' or taking away all decision power of player 2.[40] The second reason is that the match between past behavior and type is very noisy. Different types can behave in the same way. For example, both a selfish and trustworthy player 2 reciprocate on their first move, conditional on being trusted. Therefore, in order to learn about the type of a player based on past behavior, one would need many repetitions of the same game to make sure that all nodes of the game are reached several times. We wanted to rule out this type of learning. The third reason is that past experiments have shown that, in the context of games similar to the current game, the strategy to reciprocate irrespective of future interaction is typically more stable over time than other strategies (Reuben and Suetens, 2012, 2014b,a).

In order to control for γ between the treatments, we let all subjects play Imperfect and Choice. The decision taken by subjects in the role of player 2 in the above-mentioned step 1 was fixed for the two treatments. Subjects either first played 10 rounds of Imperfect and then 10 rounds of Choice, or *vice versa*. At the point that the first part (the first set of ten rounds) started, they did not have the instructions for the second part (the second set of ten rounds). However, they knew that there would be another part in the experiment.

Finally, in order to elicit preferences for disadvantageous inequality aversion of subjects who would play the role of player 1 in the trust games, these subjects made choices in a dictator game as shown in Figure 4.5 before they received the instructions for the trust game. The subjects were matched with a randomly chosen passive player, and were asked to choose the minimum payoff X above which they would choose LEFT. The

[40]For example, Camerer and Weigelt (1988), Anderhub et al. (2002) and Brandts and Figueras (2003) introduce robot trustworthy types.

minimum X was required to be between 1 and 34. This allowed us to get a quite sharp measurement of σ as included in equation 4.1 for each player 1. We calculate σ from the minimum X by solving the following indifference condition: $\sigma = \frac{\text{minimum X}-20}{18}$. These measurements allow us to study the effect of σ on choices of player 1 as discussed in section 4.2.3. Their aim is to help to interpret potential treatment effects. At the end of the experiment, we randomly drew a number between 1 and 34, that determined the outcome in the dictator game, depending on the subjects' choice of minimum X. Subjects learned about their payoffs from this part at the end of the experiment.

4.3.2. Experimental Procedures

We conducted the experiment in CentERlab at Tilburg University in May and October 2013. We used the experimental software toolkit *z-Tree* to program and conduct the experiment (Fischbacher, 2007b). Participants were recruited through an email list of students who are interested in participating in the experiments. A total number of 128 students participated in the experiment. It was explained to the participants that their earnings would depend on their own decisions and the decisions of others. The experiment lasted 1 hour including reading the instructions and payments.

The experiment covered 4 sessions of 32 participants each. In each session, subjects interacted in randomly matched pairs. We randomly rematched subjects in as large as possible sessions (32 subjects is the limit in CentERlab) rather than rematching them in (smaller) subpopulations within sessions. The first reason is that we wanted to limit the variation in proportion of trustworthy types (that is, γ) between 'populations' of subjects. The second reason is that we wanted to keep the incentives for trustees to strategically choose IN on both moves (e.g., Kandori, 1992) as low as possible.

At the beginning of each session subjects were randomly allocated a role. The roles remained fixed throughout. Participants were given instructions depending on their roles and the treatment (see section B in the Appendix for the instructions). The instructions were written in neutral wordings. All subjects went through a number of control questions after they had read the instructions. The games did not start before everyone had correctly answered the control questions.

At the end of the experiment subjects were paid in private and in cash. The payoffs in the experiment were expressed in points. Subjects were shown their total earnings in points on the screen, which equaled the sum of earnings in points over the different parts and rounds of the experiment. Earnings in Euro were determined by converting the total earnings in points into Euro at the exchange rate of 5 points= 1 Euro. The average earnings in the experiment were 16 Euro.

Table 4.1 gives an overview of the sessions.

Table 4.1: Session Summary

	Session 1	Session 2	Session 3	Session 4
Part 0	Dictator Game	Dictator Game	Dictator Game	Dictator Game
Part 1	Imperfect	Choice	Imperfect	Choice
Part 2	Choice	Imperfect	Choice	Imperfect
# Subjects	32	32	32	32

Notes: The table gives an overview of the order with which treatments are implemented and the number of subjects in each session.

4.3.3. Research Questions

Equilibrium predictions crucially depend on γ. Under 'standard' selfish preferences, subjects in the role of player 1 are predicted to choose NOT KNOW in Choice if $\gamma > \frac{1}{5}$. In this case, no difference should be observed in the first-move STAY rate of player 1 between Imperfect and Choice. In addition, if the stronger condition $\gamma > \frac{1}{3}$ holds, no differences between Choice and Imperfect should be observed in second-move STAY rates either.

If allowing for player 1 to have a utility function as in eq. (4.1), predictions additionally depend on σ. Provided that $-\frac{2}{9} < \sigma < \frac{2}{9}$, subjects in the role of player 1 are now predicted to choose NOT KNOW in Choice if $\gamma > \frac{(2-9\sigma)^2}{20-36\sigma+81\sigma^2}$. Under this assumption, no difference should be observed in the first- and second-move STAY rate of player 1 between Imperfect and Choice.

However, there are two reasons why the same player 1 may choose STAY on his first move in Imperfect, and KNOW and STOP in Choice. A first reason is that γ is not sufficiently high so as to induce him not to know in Choice, but still sufficiently high so as to induce him to STAY on his first move in Imperfect. This is the case if $\frac{1}{9} < \gamma < \frac{1}{5}$. With utility function eq. (4.1), this condition becomes $\left(\frac{2-9\sigma}{6-9\sigma}\right)^2 < \gamma < \frac{(2-9\sigma)^2}{20-36\sigma+81\sigma^2}$ (provided that $-\frac{2}{9} < \sigma < \frac{2}{9}$).[41] Interestingly, the smaller σ within interval $[-\frac{2}{9}, \frac{2}{9}]$, the larger the range of values for γ for which this condition holds.[42]

A second reason is 'strong' inequality aversion. If player 1 is strongly inequality averse so that $\sigma < -\frac{2}{9}$, in Choice he will prefer to KNOW and STOP on his first move even if γ is sufficiently high.[43] In Imperfect, however, he will still prefer to STAY on his first (and

[41] Note that if $-\frac{2}{9} < \sigma < \frac{2}{9}$, it always holds that $\frac{(2-9\sigma)^2}{20-36\sigma+81\sigma^2} < \frac{2-9\sigma}{6-9\sigma}$.

[42] $(\frac{2-9\sigma}{6-9\sigma})^2 - \frac{(2-9\sigma)^2}{20-36\sigma+81\sigma^2}$ is a negative function of σ.

[43] Previous evidence indicates that at least some subjects dislike disadvantageous inequality aversion to such an extent that $\sigma < -\frac{2}{9} \approx -0.222$. See, for example, Fehr and Schmidt (1999) and Blanco et al. (2011).

second) move if γ is sufficiently high (that is, if $\gamma > \frac{2-9\sigma}{6-9\sigma}$). If $\sigma < -\frac{2}{9}$ and $\gamma > \frac{2-9\sigma}{6-9\sigma}$ it can thus be reasonably expected that the same subject chooses STAY on both moves in Imperfect and KNOW and STOP in Choice.[44] Also, provided that $\sigma < \frac{2}{9}$, it can thus also be expected that as σ is smaller, the within-subject difference in first- and second-move STAY rates between cases where player 1 is uninformed (in Imperfect) and cases where he is informed that player 2 is selfish (in Choice) is larger.

Given that player 2 is assumed to be perfectly informed in all cases, theoretical predictions for the selfish types who play in the role of player 2 are fully driven by those for player 1. Therefore, we formulate our research questions focusing on player 1's behavior. Our main set of research questions is the following:

Question 1 *(a) Does player 1 prefer to know the type of the matched player 2 in* Choice*? (b) If so, is player 1's first- and second-move STAY rate lower in* Choice *than in* Imperfect*?*

Our second set of research questions concerns the association between σ and player 1's behavior. Under the assumption that player 1's choice in the dictator game played in the first part of the experiment measures his σ in the utility function specified in eq. (4.1), and assuming heterogeneity in the elicited σ's, we should see that player 1's choice to know the type of the matched player 2 is negatively related to his elicited σ. In addition, as argued before, it can be expected that σ is negatively related to within-subject differences in STAY rates between cases where player 1 is uninformed (cf. Imperfect) and cases in Choice where he knows player 2 is a selfish type. We formulate our second set of research questions as follows:

Question 2 *(a) Is player 1's choice to know the type of the matched player 2 in* Choice *negatively related to his elicited σ? (b) Are within-subject differences in STAY rates of player 1 between, on the one hand,* Imperfect *and, on the other hand, cases in* Choice *where he is informed about the matched player 2 being a selfish type negatively related to the elicited σ's?*

[44]Suppose, for example, that $\sigma = -\frac{2}{8}$. Player 1 then chooses KNOW and STOP in Choice. In Imperfect player 1 will choose STAY on his first (second) move if $\gamma > (\frac{17}{33})^2 \approx 0.265$ ($\gamma > \frac{17}{33} \approx 0.515$).

4.4. Experimental Results

4.4.1. Player 1's Behavior

In this section we present the main experimental results. As described in section 4.2, theoretical predictions crucially depend on the percentage of trustworthy player 2 types, namely γ. Therefore, before studying the behavior of subjects in the role of player 1, we report the observed γ (the percentage of subjects in the role of player 2 who chose IN on both moves) in the different sessions.[45] The γ's are 0.687, 0.5, 0.375, and 0.5, respectively, in Session 1 to 4.

In what follows, we report the results so as to answer the research questions in a chronological order.

4.4.1.1. Choice to KNOW

All γ's are larger than $\frac{1}{5} = 0.2$. This implies that, under the assumption that $\sigma = 0$, all subjects should choose NOT KNOW in Choice, and no difference should be observed in the STAY rates of player 1, nor in the rate at which player 2 chooses in on her first move.

We find that almost all subjects in Choice choose to KNOW. Overall, across all sessions, 90% of the subjects in the role of player 1 choose KNOW. This percentage is similar in the different sessions (varying from 82% to 92%). Furthermore, it is very similar in the two parts (87% in part 1 and 92% in part 2) so does not seem to depend on whether Choice is played before or after Imperfect. Finally, the percentage does not decrease over time. To illustrate, in the first 5 rounds it equals 87%, and in the last 5 rounds, it equals 91%.[46]

Thus we answer yes to Question 1 (a).

4.4.1.2. Behavior in Choice versus Imperfect

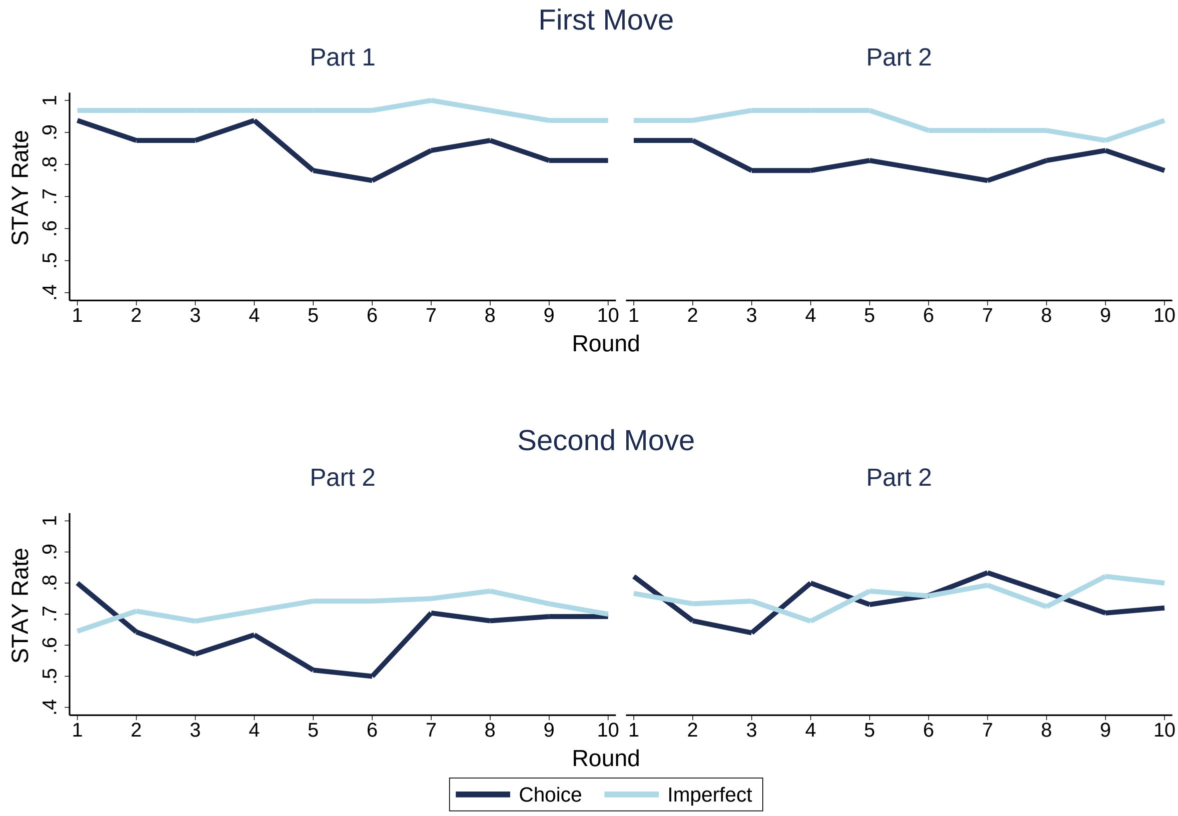
In this section, we study Question 1(b). Specifically, given that the majority of subjects in the role of player 1 prefer to KNOW in Choice, we study whether first- and second-move STAY rates are lower in Choice than in Imperfect.

Figure 4.6 shows the STAY rates of player 1 across the different rounds of the trust game. The upper panel shows the first-move STAY rate of player 1, both for part 1

[45] Since we measure the type of subjects in the role of player 2 by asking their strategy the first time they play the trust game, γ may differ between sessions.

[46] We observe an increase in the information acquisition within a session, but this increase is insignificant.

Figure 4.6: Evolution of STAY Rate of Player 1 by Treatment



Notes: The figure shows the evolution of STAY rates of subjects in the role of player 1.

(on the left-hand side) and for part 2 (on the right-hand side). The lower panel shows equivalent graphs for second-move STAY rates of player 1.

The upper panel of Figure 4.6 reveals that there is a clear difference in the first-move STAY rate of player 1 between Choice and Imperfect. Subjects in the role of player 1 choose more frequently STAY on their first move in Imperfect than in Choice. In addition, in Choice, the first-move STAY rate follows a somewhat decreasing trend, while in Imperfect it is more steady over time. There are no clear differences in patterns between Part 1 and Part 2. Across all rounds, the first-move STAY rate in Choice is 85% in Part 1 and 81% in Part 2. In Imperfect it is 97% in Part 1 and 93% in Part 2.

To test whether these differences are statistically significant, we ran 6 different specifications of regressions that include session-specific random effects. The results are shown in the upper part of Table 4.2, under 'First-move'. In all regressions the dependent variable is the percentage of times subjects in the role of player 1 play STAY on their first move. In specification (1) we include as an independent variable a treatment dummy. In specification (2), next to the treatment dummy, we control for the ordering of the treatments by including a dummy that refers to Imperfect being played in Part 1. The estimated effect is 0.118 in both specifications and is statistically significant. The first-move STAY rate of player 1 is thus 11.8 percentage points higher in Imperfect than in Choice. In specifications (3) and (4), we separately estimate the treatment effect for Part 1 and Part 2 (in fact, these are between-subjects treatment effects). The estimated marginal effects are 0.115 ($p = 0.001$) and 0.121 ($p = 0.217$), respectively. In specifications (5) and (6) we report the within-subjects treatment effects. Specification (5) uses data from sessions where Choice is played in Part 1 and Imperfect in Part 2, and *vice versa* in specification (6). The estimated marginal effect in (5) is 0.081 ($p = 0.0332$). The estimated marginal effect in (6) is 0.156 ($p = 0.001$). We conclude that both between and within subjects, the first-move STAY rate is between 8.1 and 15.6 percentage points higher (significant) in Imperfect than in Choice.

Next, we focus on second-move STAY rates, shown in the lower panel of Figure 4.6. The figure shows that treatment effects are smaller than with respect to first-move STAY rates. Across all rounds, the second-move STAY rate in Choice is 65% in Part 1 and 75% in Part 2. In Imperfect it is 71% in Part 1 and 75% in Part 2. As shown in the lower part of Table 4.2, the treatment difference is weakly significant in Part 1, with the STAY rate being higher in Imperfect than in Choice, but in Part 2 the difference disappears. Our answer to Question 1(b) is thus partly yes — with respect to player 1's first move — and partly no — with respect to player 1's second move.

In Choice, aggregate trust rates hide quite a lot of information. Indeed, the behavior of an informed player 1 is expected to crucially depend on whether he is matched with a trustworthy or selfish player 2. We now study differences in STAY rates of player 1 on his first and second move depending on whether the matched player 2 is a trustworthy or a selfish type.

Figure 4.7 shows the evolution of STAY rates of subjects in the role of player 1 who choose to KNOW. The upper panel shows the first-move STAY rate and the lower panel the second-move STAY rate (with Part 1 on the left-hand side and Part 2 on the right-hand side). As can be seen in the figure, first- and second-move STAY rates strongly depend on the type of the matched player 2. If player 2 is a trustworthy type, subjects in the role of player 1 are much more inclined to STAY on their first and second move than if player 2 is a selfish type. The first-move STAY rate is almost 100% throughout the whole experiment when player 1 knows that the matched player 2 is a trustworthy type. In the case where player 1 knows that the matched player 2 is a selfish type, the first-move STAY rate drops to 70% in Part 1 and 56% in Part 2. In addition, the first-move STAY rate of a player 1 matched with a selfish player 2 exhibits a decreasing trend over time. Moreover, the second-move STAY rate is almost 90% throughout the whole experiment when player 1 knows that the matched player 2 is a trustworthy type, while it drops to 20% when the matched player 2 is a selfish type.

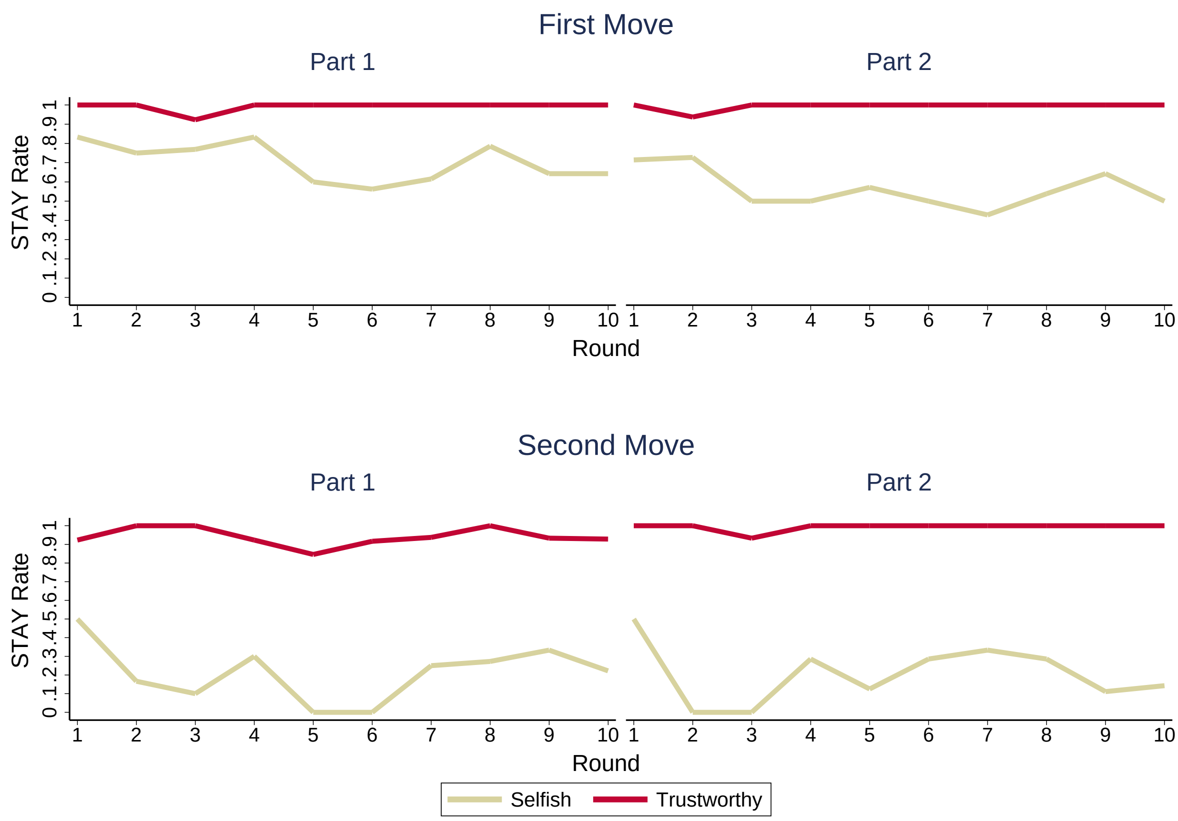
Table 4.2: Treatment Effects on the STAY Rate of Player 1

	(1)	(2)	(3)	(4)	(5)	(6)
			First move			
Imperfect	0.118***	0.118***	0.115**	0.121	0.081*	0.156***
	(0.026)	(0.026)	(0.034)	(0.098)	(0.038)	(0.036)
Imperfect in Part 1		−0.003				
		(0.060)				
Constant	0.830***	0.831***	0.850***	0.810***	0.850***	0.810***
	(0.028)	(0.045)	(0.024)	(0.069)	(0.027)	(0.062)
# Observations	128	128	64	64	64	64
R^2	0.129	0.129	0.154	0.115	0.068	0.202
			Second move			
Imperfect	0.017	0.017	0.054	−0.020	0.083*	−0.050
	(0.043)	(0.043)	(0.062)	(0.140)	(0.063)	(0.059)
Imperfect in Part 1		0.037				
		(0.094)				
Constant	0.708***	0.689***	0.655***	0.760***	0.655***	0.760***
	(0.046)	(0.070)	(0.044)	(0.099)	(0.069)	(0.078)
# Observations	128	128	64	64	64	64
R^2	0.001	0.007	0.012	0.002	0.026	0.011

Notes: The table reports estimations (standard errors in parentheses) from linear regressions that include session-specific random effects. The dependent variable is the percentage of times player 1 chooses STAY. The independent variable Imperfect is a treatment dummy. *** (**) [*] indicate that the estimated coefficient is significant at the 1% (5%) [10%] level. Specification (1) is based on all observations; (2) is based on all observations, and includes a dummy which is equal to 1 if Imperfect is played in Part 1 and Choice is played in Part 2; (3) is based on the observations in Part 1; (4) is based on the observations in Part 2; (5) is based on the observations where Choice is played in Part 1 and Imperfect is played in Part 2; and (6) is based on the observations where Imperfect is played in Part 1 and Choice is played in Part 2.

Being perfectly informed about player 2's type makes player 1 choose STAY on his second move in almost 100% of the cases if player 2 is a trustworthy type and in almost 0% of the cases she is a selfish type. On his first move, however, player 1 still plays STAY in a substantial fraction of the cases (between 56% and 70% depending on the part) even if player 2 is a selfish type. This suggests that there may be other reasons for player 1 to choose STAY on his first move than those discussed in the theoretical model. One of these might be that (player 1 believes that) sufficiently many subjects in the role of player 1 choose STAY on both of their moves because they have a preference to maximize total efficiency or to be trustful. If this is the case, a selfish player 1 has

Figure 4.7: Evolution of STAY Rate of Player 1 in Choice



Notes: The figure shows the evolution of STAY rates of subjects in the role of player 1 within the Choice treatment.

strategic reasons to choose STAY on his first move. That is, if he can make player 2 believe he is a 'trustful' person, so that player 2 would choose IN on her first move in the hope that player 1 chooses STAY on his second move, player 1 will be better off if he chooses STAY on his first move than when choosing STOP.

Table 4.3 reports results of regressions with session-specific random effects, using data from Choice where the first- or second-move STAY rate of subjects in the role of player 1 (upper and lower part, respectively) is regressed on a treatment dummy. In all instances, the treatment effect is large and significant. Overall, as shown in (1), a matched player 2 being a trustworthy type increases the probability of player 1 choosing STAY on his first move by 37% ($p < 0.001$), and by 72% ($p < 0.001$) on his second move.

4.4.1.3. Elicited σ and Player 1's Behavior

The above results suggest that for a substantial part of the subjects $\sigma < 0$. The reason is that in all sessions, γ is sufficiently high so that if subjects in the role of player 1 would be selfish, they would prefer to be uninformed in Choice so that no difference is predicted between Choice and Imperfect. In this section, we study more closely the relation between

Table 4.3: Effect of Player 2's Type on Informed Player 1's STAY Rate in Choice

	(1)	(2)	(3)	(4)	(5)	(6)
			First move			
Trustworthy	0.367***	0.367***	0.296***	0.435***	0.296***	0.435***
	(0.045)	(0.046)	(0.059)	(0.069)	(0.059)	(0.069)
Imp in Part 1		−0.066				
		(0.046)				
Constant	0.628***	0.662***	0.698***	0.561***	0.698***	0.561***
	(0.037)	(0.041)	(0.042)	(0.049)	(0.042)	(0.049)
# Observations	122	122	60	62	60	62
R^2	0.347	0.358	0.302	0.398	0.303	0.398
			Second move			
Trustworthy	0.713***	0.712***	0.729***	0.696***	0.729***	0.697***
	(0.050)	(0.050)	(0.065)	(0.076)	(0.065)	(0.076)
Imp in Part 1		0.066				
		(0.050)				
Constant	0.257***	0.224***	0.215***	0.299***	0.215***	0.299***
	(0.039)	(0.044)	(0.048)	(0.056)	(0.048)	(0.056)
# Observations	116	116	58	58	58	58
R^2	0.638	0.644	0.686	0.600	0.687	0.600

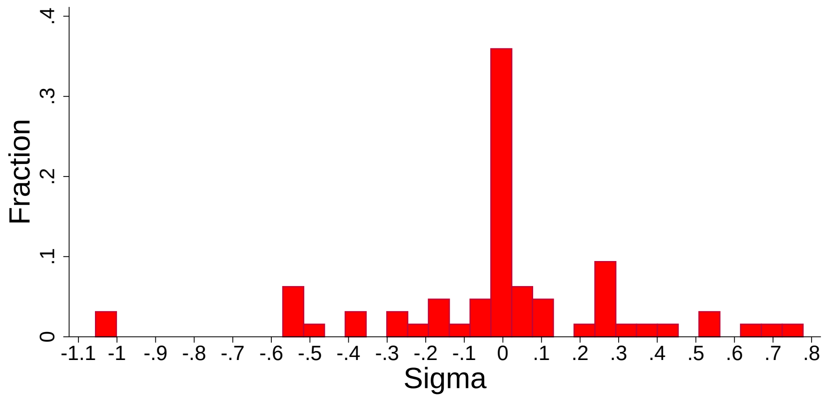
Notes: The table reports estimations (standard errors in parentheses) from linear regressions that include session-specific random effects, based on the observations where player 1 chooses to KNOW in Choice. The dependent variable is the percentage of times player 1 choose STAY. *** (**) [*] indicate that the estimated coefficient is significant at the 1% (5%) [10%] level. For an explanation of the 6 specifications, see note in Table 4.2.

σ elicited in the dictator game played in the first part of the experiment, and behavior of player 1.

Obviously, this exercise is only relevant if there exists heterogeneity in σ across the subjects in the first place. To illustrate this is the case, we refer to Figure 4.8, which shows the distribution of σ. Figure 4.8 shows that the mode is $\sigma = 0$; about 35% of the subjects have $\sigma = 0$. This means that the modal subject in the role of player 1 is a selfish payoff maximizer. The other subjects either have a negative σ, so dislike being behind in payoff terms, or a positive σ, so have a preference for efficiency.

We now proceed to answering Question 2 (a). To do so, we compute the correlation between the percentage of times subjects in the role of player 1 in Choice choose to obtain information about the matched player 2 (i.e., choose KNOW), and the elicited σ for that subject. We find a (weakly) significant correlation of -0.221 ($p = 0.079$). This is (weak)

Figure 4.8: Distribution of σ



Notes: The figure shows the distribution of σ elicited in the dictator game of 64 subjects in the role of player 1. It is calculated as follows: $\sigma = \frac{\text{minimum X}-20}{18}$.

Table 4.4: Within-Subject Difference in STAY Rates and σ's

	First Move	Second Move
Constant	0.237 (0.058)***	0.249 (0.071)***
σ	-0.551 (0.196)***	0.249 (0.304)
# Observations	47	41
R^2	0.150	0.017

Notes: The table reports estimations (standard errors in parentheses) from linear regressions at the subject level. The dependent variable is the aggregate STAY rate in Imperfect minus the aggregate STAY rate in Choice conditional on being informed that player 2 is selfish, and conditional on $\sigma < \frac{2}{9}$. *** (**) [*] indicate that the estimated coefficient is significant at the 1% (5%) [10%] level.

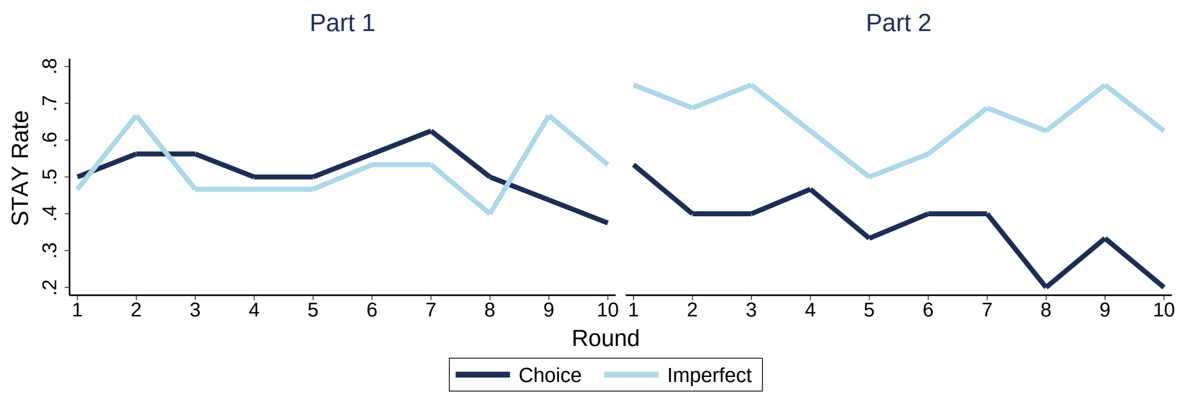
evidence in favor of σ being negatively related to player 1's willingness to KNOW the type of the matched player 2. The answer to Question 2 (a) is thus (weakly) yes.

Next, we study Question 2 (b). To do so, we use the treatment variation within subjects to study whether the differences in STAY rates between Imperfect and Choice are related to σ. Specifically, we study the relation between the elicited σ's and the within-player 1 difference in the percentage of times he chooses STAY between cases where he is uninformed (in Imperfect) and cases where he is informed about the matched player 2 being a selfish type (in Choice). Results from regressions are in Table 4.4.[47]

The table shows that the within-subject difference in first-move STAY rates between being uninformed and being informed that the matched player 2 is a selfish type is significantly positive. In addition, the difference is significantly higher, the lower is σ,

[47] A similar exercise can be done within Choice, namely between cases where player 1 chose to stay uninformed versus informed about player 2 being a selfish trustee. Given that player 1 almost always chose to be informed, we refrained from performing such exercise.

Figure 4.9: Evolution of First-Move IN Rate of Player 2



Notes: The figure shows the evolution of first-move IN rates of 31 subjects in the role of player 2 who do not choose IN on their second move (so selfish types).

provided that $\sigma < \frac{2}{9}$. This is in line with the theoretical predictions. For the second-move STAY rates we find that within-subject differences are positive, as expected, but are not significantly associated with σ. Our answer to Question 2(b) is thus partly yes — with respect to player 1's first move — and partly no — with respect to player 1's second move.

4.4.2. Player 2's Behavior

In this section we provide a description of player 2's behavior. We first study the rate at which subjects in the role of player 2 choose IN. We focus on the so-called selfish types, namely, on those subjects who did not choose IN on both moves (or, equivalently, who chose OUT on their second move). Figure 4.9 shows the first-move IN rates of selfish players 2 by round in Part 1 (on the left-hand side) and Part 2 (on the right-hand side).

Figure 4.9 shows that there is no clear difference in the first-move IN rate between the treatments in Part 1. It is approximately 50% in both treatments and does not evolve in a clear direction over time. In Part 2, however, the first-move IN rate is clearly higher in Imperfect than in Choice. On average, subjects in the role of (a selfish) player 2 choose IN on their first move 65% of the time in Imperfect, and only 40% of the time in Choice. In addition, in Choice this rate decreases over time.

Table 4.5 reports the results from regressions with session-specific random effects, using the same 6 specifications as in the regressions for player 1, but now using the first-move IN rate of player 2 as a dependent variable.

As can be seen in the table, the treatment effect is positive, as would be expected by theory, but it is not statistically significant. Overall, the first-move IN rate is estimated to be 7.1 percentage points higher in Imperfect than in Choice ($p = 0.233$ in both (1) and

Table 4.5: Treatment Effects on First-Move IN Rate of Player 2

	(1)	(2)	(3)	(4)	(5)	(6)
Imperfect	0.071	0.071	0.018	0.125	0.071	0.072
	(0.060)	(0.060)	(0.103)	(0.150)	(0.085)	(0.084)
Imp in Part 1		−0.053				
		(0.126)				
Constant	0.730***	0.756***	0.756***	0.703***	0.756***	0.703***
	(0.062)	(0.094)	(0.073)	(0.106)	(0.061)	(0.130)
# Observations	128	128	64	64	64	64
R^2	0.011	0.016	0.001	0.029	0.011	0.010

Notes: The table reports estimations (standard errors in parentheses) from linear regressions that include session-specific random effects. The dependent variable is the percentage of times player 2 chooses IN. The independent variable Imperfect is a treatment dummy. *** (**) [*] indicate that the estimated coefficient is significant at the 1% (5%) [10%] level. For an explanation of the 6 specifications, see note in Table 4.2.

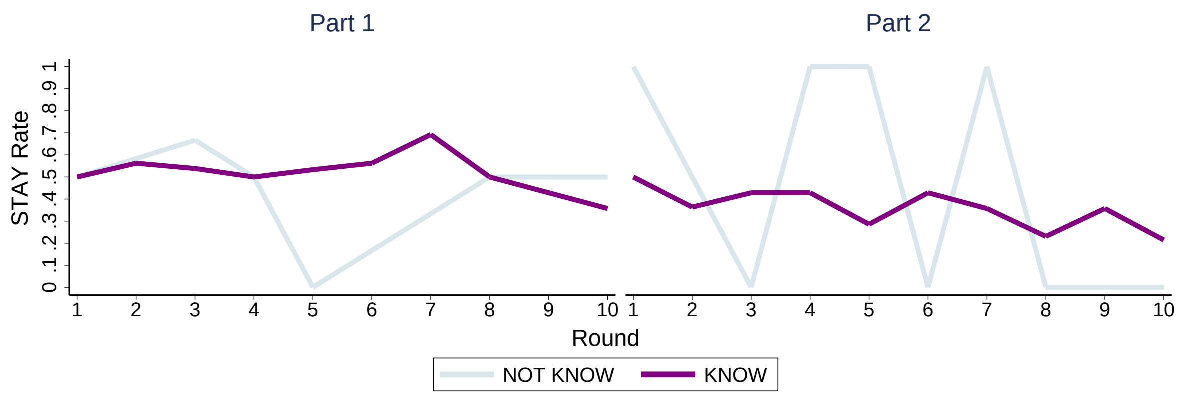
(2)). Although the magnitude of the marginal effect is larger in Part 2 (cf. column (4) in the table), the treatment effect not significant either ($p = 0.403$).

Next, we study whether the behavior of (a selfish) player 2 in Choice depends on whether the matched player 1 knows about her type. Specifically, we focus on how the first-move IN rate depends on whether player 1 knows about the type of the matched player 2 or not. Figure 4.10 depicts the evolution of this first-move IN rate (Part 1 in the left-hand panel and Part 2 in the right-hand panel). We should note that because so few subjects in the role of player 1 choose NOT KNOW, the lines related to NOT KNOW are based on very few data points, which may explain why they are fluctuating to a large extent. No clear difference can be seen in the figure between cases where player 1 knows about the type of player 2 and cases where he does not know. As shown in Table 4.6, which reports results from regressions with session-specific random effects, although in the theoretically predicted direction, the effect is not significant in a statistical sense either.

4.5. Discussion

In summary, we find that in our experimental twice-repeated trust games, trustors typically choose to be informed about the matched trustee's type. We also find that being uninformed leads to significantly more trust today (on the first move) than being informed that the matched trustee is selfish. No significant effect appears on tomorrow's (second-move) trust rate, though. The consequence is that trustors are generally not

Figure 4.10: Evolution of Player 2's First-Move IN Rate in Choice



Notes: The figure shows the evolution of first-move IN rates of subjects in the role of player 2 who do not choose IN on their second move (so selfish types).

Table 4.6: Effect of Knowing on Player 2's First-Move IN Rate

	(1)	(2)	(3)	(4)	(5)	(6)
Know	−0.059	−0.058	0.014	−0.136	0.014	−0.136
	(0.113)	(0.114)	(0.155)	(0.172)	(0.155)	(0.172)
Imp in Part 1		−0.084				
		(0.147)				
Constant	0.501***	0.544***	0.500***	0.500***	0.500***	0.500***
	(0.089)	(0.120)	(0.117)	(0.133)	(0.117)	(0.133)
# Observations	53	53	28	25	28	25
R^2	0.005	0.017	0.001	0.026	0.001	0.026

Notes: The table reports estimations (standard errors in parentheses) from linear regressions that include session-specific random effects, based on the observations where player 2 is a selfish type. The dependent variable is the percentage of times player 2 chooses IN. *** (**) [*] indicate that the estimated coefficient is significant at the 1% (5%) [10%] level. For an explanation of the 6 specifications, see note in Table 4.2.

worse off when informed than when uninformed. To illustrate, the average material payoff if the trustee is a trustworthy type is 31.65 when informed and is 28.50 when not informed. If the trustee is a selfish type, the average material payoff of the trustor is 16.74 when informed and 16.45 when not informed.[48]

Part of the qualitative findings can be organized by allowing for non-standard preferences of the trustor in a model where the trustor is uninformed about the trustee's

[48]The average utility à la Charness and Rabin (2002) is very similar to the material payoffs. We calculate the utilities by using the elicited σ values and the utility function presented in Equation (1). The average utility if the trustee is a trustworthy type is 31.65 when informed and is 28.50 when not informed. If the trustee is a selfish type, the average material payoff of the trustor is 16.95 when informed and 17.34 when not informed.

type. Specifically, we find that one reason why trustors prefer to know the type of the matched trustee seems to be that they subtract utility from disliking getting a lower payoff than others. For one, we find that the percentage of times trustors choose to be informed is positively associated with the extent of disliking getting a lower payoff, elicited in a dictator game played before the trust games. Second, we find that this preference parameter helps to organize how trust rates differ within trustors between being informed about the matched trustee being a selfish type and being uninformed. We cannot rule out alternative explanations, however. Regret aversion or betrayal aversion, for example, may also lead the trustor to prefer to know the type of the trustee (see Dominguez-Martinez et al. (2014) and Bohnet and Zeckhauser (2004)). In our context, regret or betrayal averse agents derive disutility from finding out their decisions are not the ones they would have taken if they had been informed about the trustee's type. This would mean that a trustful trustor finding out the trustee did not reciprocate suffers more than what we have assumed in our model.

Some of the findings do not fit within the proposed theoretical framework, though. One of these is already mentioned, namely that there is no significant treatment effect on second-move trust rates. Second, first-move trust rates of trustors who are informed about the matched trustee being a selfish type are well above 0, namely approximately 50%. Third, no or only small differences are observed in first-move reciprocation rates of trustees between Choice and Imperfect. We speculate that the reason why the proposed model does not capture these findings is that it does not allow for the trustee being uninformed about the type of the trustor. Particularly, by trusting on his first move, trustor may induce the trustee to believe he is a 'trustful' of efficiency-maximizing person. The trustee then reciprocates on her first move in the hope that the trustor trusts again on his second move. This mechanism may give incentives to trustors to trust (strategically) on their first move even when being perfectly informed that the matched trustee is selfish. In future work, it would be interesting to develop models that allow for such uncertainty.

Appendix

4.A. Trustors with Social Preferences

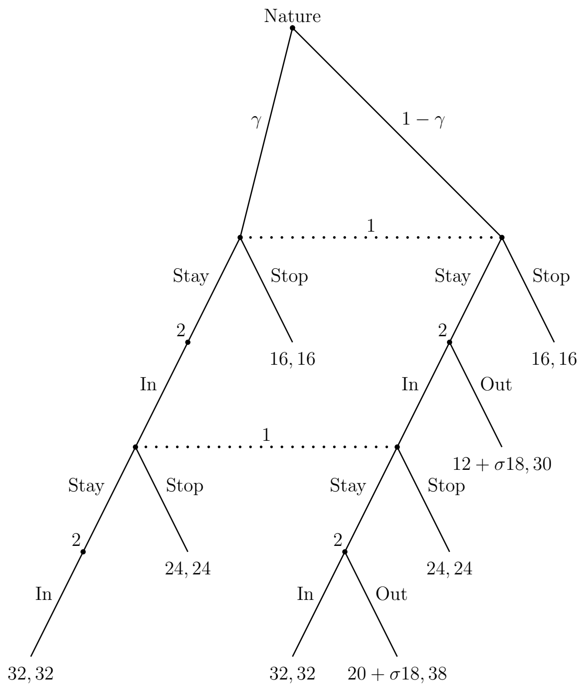
4.A.1. Imperfect Information Game

In this section we analyze the game played in the Imperfect treatment, under the assumption that player 1 has the following utility function:

$$U(x_1, x_2) = x_1 + \sigma(x_2 - x_1), \tag{4.2}$$

where x_1 and x_1 are the material payoffs for player 1 and player 2. Figure 4.11 illustrates the utilities of the Imperfect game when the trustor is inequality averse. In the following

Figure 4.11: Imperfect



we solve the game.

In the last period, a trustworthy player 2 chooses IN and a selfish player 2 chooses OUT by definition. Player 1 chooses STAY in the last information set if his expected payoff from choosing STAY is larger than his expected payoff from choosing STOP. Note that if $\sigma > \frac{2}{9}$ player 1 is better off by playing STAY on his last move irrespective of player 2' move. This is because having a $\sigma > \frac{2}{9}$, player 1 prefers himself trusting on his last move and player 2 not reciprocating after that to himself not trusting on this last move. If, however,$\sigma < \frac{2}{9}$, player 1, observing the first move of player 2, updates his belief about player 2's type using Bayesian updating. In the following we first discuss the case where $\sigma < \frac{2}{9}$ and then the case where $\sigma > \frac{2}{9}$ separately.

First assume that $\sigma < \frac{2}{9}$. Denote the updated belief after the first move of player 2 by μ_2. Player 1 chooses STAY on his second move if $\mu_2 32 + (1 - \mu_2)(20 + \sigma 18) > 24$. That is, player 1 plays STAY on his second move if $\mu_2 > \frac{2-9\sigma}{6-9\sigma}$. [49]

We first consider the case when $\gamma > \frac{2-9\sigma}{6-9\sigma}$. Applying Bayes' Rule $\mu_2 > \gamma > \frac{2-9\sigma}{6-9\sigma}$. In this case player 1 chooses STAY on his second move. Player 2 plays IN on her first move if she is trustworthy. If player 2 is selfish, then she also chooses IN on her first move since $38 > 30$. Player 1 chooses STAY in the first information set if his expected payoff from choosing STAY is larger than his expected payoff from choosing STOP, that is if $\gamma 32 + (1-\gamma)(20+\sigma 18) > 16$. This last inequality reduces to $\gamma > \frac{-4}{12-18\sigma}$. This inequality always holds with $\sigma < \frac{2}{9}$, since γ can not be negative. Thus player 1 chooses STAY on his first move.

We then consider the case when $\gamma < \frac{2-9\sigma}{6-9\sigma}$. We first show that the equilibrium is in mixed strategies in this case. Let m be the probability that player 1 chooses STAY on his second move in PG and p be the probability that a selfish player 2 chooses IN on her first move. Player 1 updates his belief according to Bayes' Rule. His updated belief given that player 2 chooses IN on her first move is given by $\mu_2 = \frac{\gamma}{\gamma+p(1-\gamma)}$. We first show that p cannot take the values 0 and 1 (we follow Anderhub et al. (2002)). Suppose that $p = 0$, then by Bayesian updating, $\mu_2 = \frac{\gamma}{\gamma+0(1-\gamma)} = 1$. This implies that player 1 optimally plays STAY ($(1 - m) = 0$), in which case player 2 would optimally choose IN (since $38 > 30$) with probability $p = 1$ contradicting the initial supposition. Thus $p > 0$ should be satisfied. Suppose that $p = 1$. Then by Bayesian updating, $\mu_2 = \frac{\gamma}{\gamma+1(1-\gamma)} = \gamma < \frac{2-9\sigma}{6-9\sigma}$. This implies that $m = 0$. But then player 2 would chose OUT on her first move since $30 > 24$, which contradicts with the initial supposition that $p = 1$. Thus $0 < p < 1$ should hold. That is, player 2 uses a mixed strategy on her first move. In this case, player 1 should also use a mixed strategy on his second move as otherwise player 2 would optimally play a pure strategy on her first move as well. Therefore, we have proven that $0 < p < 1$ and $0 < m < 1$ when $\gamma < \frac{2-9\sigma}{6-9\sigma}$.

Next we calculate the equilibrium mixing probabilities. Player 1, on his second move, plays STAY with probability m and plays STOP with probability $(1 - m)$. He chooses m such that a selfish player 2 is indifferent on her first move. Thus we have $m38 + (1 - m)24 = 30$, yielding $m = \frac{3}{7}$. Player 2, on her first move, plays IN with probability p and

[49]Note that if $\sigma > \frac{2}{9}$ then $20 + \sigma 18 > 32$. In this case, player 1 chooses STAY on his second move.

plays OUT with probability $(1-p)$. He chooses p such that player 1 is indifferent on his second move, that is $\mu_2 = \frac{\gamma}{\gamma+p(1-\gamma)} = \frac{2-9\sigma}{6-9\sigma}$, yielding $p = \frac{4\gamma}{(1-\gamma)(2-9\sigma)}$. Player 1, on his first move, plays STAY if his expected payoff from doing so is larger than his expected payoff from choosing STOP. That is, if $\gamma[m32 + (1-m)24] + (1-\gamma)[p(m(20+\sigma18) + (1-m)24) + (1-p)(12+\sigma18)] > 16$. Substituting $m = \frac{3}{7}$ and $p = \frac{4\gamma}{(1-\gamma)(2-9\sigma)}$, we get $\gamma > \frac{4-36\sigma+81\sigma^2}{36-108\sigma+81\sigma^2} = (\frac{2-9\sigma}{6-9\sigma})^2$. Thus player 1 chooses STAY if $\frac{2-9\sigma}{6-9\sigma} > \gamma > (\frac{2-9\sigma}{6-9\sigma})^2$ and chooses STOP if $\gamma < (\frac{2-9\sigma}{6-9\sigma})^2$. The equilibrium predictions are summarized as follows:

Prediction Imperfect, $\sigma < \frac{2}{9}$. *Let player 2 be a trustworthy type with probability γ and a selfish type with probability $1-\gamma$. The theoretical predictions of the Imperfect game are:*

(a) $\gamma > \frac{2-9\sigma}{6-9\sigma}$: Player 1 plays STAY on both moves. Beliefs in the first and second information set are $\mu_1 = \gamma$ and $\mu_2 = \gamma$, respectively. A trustworthy player 2 plays IN on both moves and a selfish player 2 plays IN on her first move and OUT on her second move.

(b) $(\frac{2-9\sigma}{6-9\sigma})^2 < \gamma < \frac{2-9\sigma}{6-9\sigma}$: Player 1 plays STAY on his first move and plays STAY with probability $\frac{3}{7}$ and STOP with probability $\frac{4}{7}$ on his second move. Beliefs in the first and second information set are $\mu_1 = \gamma$ and $\mu_2 = \frac{2-9\sigma}{6-9\sigma}$, respectively. A trustworthy player 2 plays IN on both moves. A selfish player 2 plays IN with probability $\frac{4\gamma}{(1-\gamma)(2-9\sigma)}$ and OUT with probability $1-\frac{4\gamma}{(1-\gamma)(2-9\sigma)}$ on her first move and plays OUT on her second move.

(c) $\gamma < (\frac{2-9\sigma}{6-9\sigma})^2$: Player 1 plays STOP on his first move and plays STAY with probability $\frac{3}{7}$ and STOP with probability $\frac{4}{7}$ on his second move. Beliefs in the first and second information set are $\mu_1 = \gamma$ and $\mu_2 = \frac{2-9\sigma}{6-9\sigma}$, respectively. A trustworthy player 2 plays IN on both moves. A selfish player 2 plays IN with probability $\frac{4\gamma}{(1-\gamma)(2-9\sigma)}$ and OUT with probability $1-\frac{4\gamma}{(1-\gamma)(2-9\sigma)}$ on her first move and OUT on her second move.

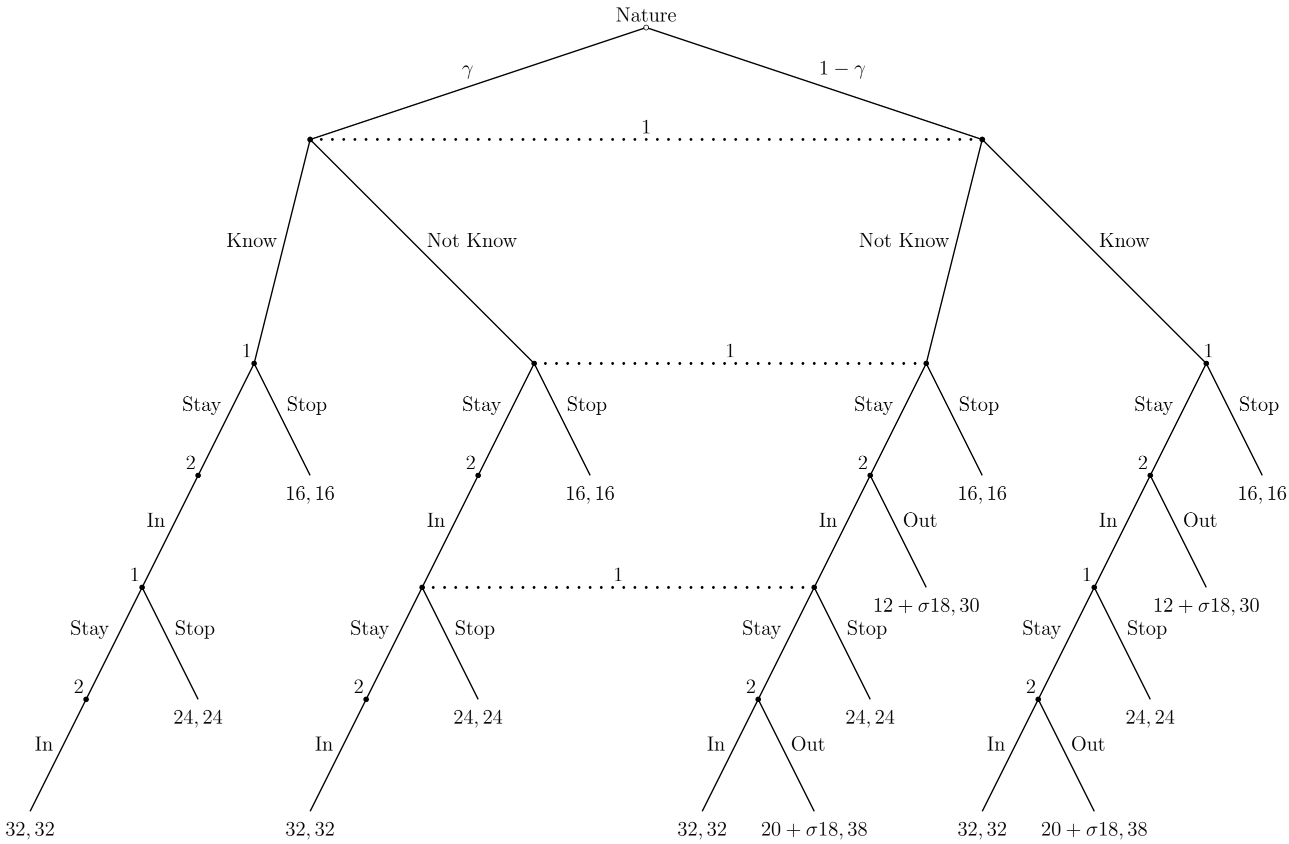
Second assume that $\sigma > \frac{2}{9}$. In this case player 1 chooses STAY on his last move irrespective of player 2's type. This is because he is better off if he trusts on his last move even if player 2 does not reciprocate this trust whenever $\sigma > \frac{2}{9}$. A trustworthy player 2 chooses IN on his first move by definition. A selfish player 2 chooses IN on his first move as well since $38 > 30$. Player 1, on his first move, chooses STAY if $16 < \gamma 32 + (1-\gamma)(20+\sigma 18)$, which always true. The equilibrium predictions are summarized as follows:

Prediction Imperfect, $\sigma > \frac{2}{9}$. *Let player 2 be a trustworthy type with probability γ and a selfish type with probability $1-\gamma$. The theoretical predictions of the* Choice *game* *are: Player 1 plays STAY on his both moves. A trustworthy player 2 plays IN on his both moves and a selfish player 2 plays IN on her first move and OUT on her second move.*

4.A.2. Choice game

In this section we reconsider the game illustrated in Choice treatment, under the assumption that the utility of player 1 is given by the equation 4.2, illustrated in Figure 4.12.

Figure 4.12: Choice with inequality averse trustors



Before calculating the optimal choice of player 1, that is, whether player 1 chooses to KNOW or NOT KNOW the type of player 2, we solve the subgames depending on player 1's choice in period 2.

4.A.2.1. *Player 1 chooses KNOW*

Suppose that player 1 chooses KNOW in period 2 and we solve the game by backward induction. In the last period of the game, a trustworthy player 2 chooses IN and a selfish player 2 chooses OUT. Player 1, on his second move, chooses STAY if player 2 is trustworthy since $32 > 24$. Player 1, on his second move, chooses STAY if player 2 is selfish only when $20 + \sigma 18 > 24$. This is the case whenever $\sigma > \frac{2}{9}$. First consider the case when $\sigma > \frac{2}{9}$, so that player 1 chooses STAY on his second move when player 2 is selfish. A trustworthy player 2 chooses IN on his first move by definition. A selfish player 2 chooses IN on her first move as $38 > 30$. Player 1, on his first move , chooses STAY if he is matched with a trustworthy player 2 since $32 > 16$. Player 1, on his first move, chooses STAY if he is matched with a selfish player 2 only when $20 + \sigma 18 > 16$. This inequality always holds since $\sigma > \frac{2}{9}$. Second consider the case when $\sigma < \frac{2}{9}$, so that player 1 chooses STOP on his second move when player 2 is selfish. A trustworthy player 2 chooses IN on his first move by definition. A selfish player 2 chooses OUT on her first move as $30 > 24$. Player 1, on his first move, chooses STAY if $16 < 12 + \sigma 18$. As $\sigma < \frac{2}{9}$, this inequality does not hold, implying player 1 chooses STOP on his first move. So, in summary, given that player 1 chooses KNOW in period 2, we have two different predictions depending on the value of σ. In summary if $\sigma < \frac{2}{9}$: A trustworthy player 2 will choose IN on both moves and player 1 chooses STAY on both moves in equilibrium. A selfish player 2 chooses OUT on both moves and player 1 matched with a selfish player 2 chooses STOP on both moves in equilibrium. If $\sigma > \frac{2}{9}$: A trustworthy player 2 will choose IN on both moves and player 1 chooses STAY on both moves in equilibrium. A selfish player 2 chooses IN on his first move, OUT on his second move and player 1 matched with a selfish player 2 chooses STAY on both moves in equilibrium.

4.A.2.2. *Player 1 chooses NOT KNOW*

When player 1 chooses NOT KNOW, we are in the Imperfect game. We refer to section 4.2.1 for the calculation of the equilibrium in this subgame.

4.A.2.3. *Choice to KNOW or NOT KNOW*

We have shown that the predictions for the trust game in the cases where $\sigma < \frac{2}{9}$ and $\sigma > \frac{2}{9}$ are different, as explained in the previous section. First we discuss the case where $\sigma < \frac{2}{9}$ and we second discuss the case where $\sigma > \frac{2}{9}$.

We now calculate the optimal choice of player 1 in period 2, when $\sigma < \frac{2}{9}$. Player 1 chooses KNOW if his expected payoff from playing KNOW is larger than his expected payoff from NOT KNOW. We consider the 3 ranges for γ that are relevant given that player 1 does NOT KNOW (see 4.A.1): $\gamma > \frac{2-9\sigma}{6-9\sigma}$, $\frac{2-9\sigma}{6-9\sigma} > \gamma > (\frac{2-9\sigma}{6-9\sigma})^2$ and $\gamma < (\frac{2-9\sigma}{6-9\sigma})^2$.

First, we consider the case where $\gamma > \frac{2-9\sigma}{6-9\sigma}$. Player 1 chooses KNOW in this case if

$$\gamma 32 + (1-\gamma)16 > \gamma 32 + (1-\gamma)(20 + \sigma 18). \tag{4.3}$$

This inequality holds only when $\sigma < -\frac{2}{9}$. Thus player 1 chooses KNOW if $\sigma < -\frac{2}{9}$ and chooses NOT KNOW if $\sigma > -\frac{2}{9}$.

Second, we consider the case where $(\frac{2-9\sigma}{6-9\sigma})^2 < \gamma < \frac{2-9\sigma}{6-9\sigma}$. Player 1 chooses KNOW in this case if

$$\gamma 32 + (1-\gamma)16 > \gamma[R32 + (1-m)24] + (1-\gamma)[p(m(20+\sigma 18) + (1-m)24) + (1-p)(12+\sigma 18)]. \tag{4.4}$$

Plugging in the equilibrium values for m and p derived above, this inequality holds only when $\gamma < \frac{4-36\sigma+81\sigma^2}{20-36\sigma+81\sigma^2}$. It is important to note that $\frac{4-36\sigma+81\sigma^2}{20-36\sigma+81\sigma^2} < \frac{2-9\sigma}{6-9\sigma}$ only if $\sigma > -\frac{2}{9}$ and $\frac{4-36\sigma+81\sigma^2}{20-36\sigma+81\sigma^2} > \frac{2-9\sigma}{6-9\sigma}$ if $\sigma < -\frac{2}{9}$. In summary, whenever $(\frac{2-9\sigma}{6-9\sigma})^2 < \gamma < \frac{2-9\sigma}{6-9\sigma}$ the optimal decision of player 1 in period 2 is summarized as follows: if $\sigma < -\frac{2}{9}$ player 1 chooses KNOW for $(\frac{2-9\sigma}{6-9\sigma})^2 > \gamma > \frac{4-36\sigma+81\sigma^2}{20-36\sigma+81\sigma^2}$ and if $\sigma > -\frac{2}{9}$ player 1 chooses KNOW whenever $(\frac{2-9\sigma}{6-9\sigma})^2 < \gamma < \frac{4-36\sigma+81\sigma^2}{20-36\sigma+81\sigma^2}$ and chooses NOT KNOW whenever $\frac{4-36\sigma+81\sigma^2}{20-36\sigma+81\sigma^2} > \gamma > \frac{2-9\sigma}{6-9\sigma}$. Player 1 is indifferent between choosing KNOW and NOT KNOW whenever $\gamma = \frac{4-36\sigma+81\sigma^2}{20-36\sigma+81\sigma^2}$.

Last, we consider the case where $\gamma < (\frac{2-9\sigma}{6-9\sigma})^2$. Player 1 chooses KNOW in this case if

$$\gamma 32 + (1-\gamma)16 > 16.$$

This inequality always holds, thus player 1 chooses KNOW whenever $\gamma < (\frac{2-9\sigma}{6-9\sigma})^2$.

We summarize the equilibrium predictions as follows:

Prediction Choice, $\sigma < \frac{2}{9}$. *Let player 2 be a trustworthy type with probability γ and a selfish type with probability $1-\gamma$. The theoretical predictions of the* Choice *game are:*

(a) $\gamma > \frac{2-9\sigma}{6-9\sigma}$: Player 1 plays NOT KNOW if $\sigma > -\frac{2}{9}$ and KNOW if $\sigma < -\frac{2}{9}$. Player 1 plays STAY on both moves if he plays NOT KNOW. Beliefs in the first and second information set are $\mu_1 = \gamma$ and $\mu_2 = \gamma$, respectively. When player 1 plays KNOW, then he plays STAY on both moves when player 2 is a trustworthy type and STOP on both moves when player 2 is a selfish type. A trustworthy player 2 plays IN on both moves when player 1 plays KNOW or NOT KNOW. A selfish player 2 plays IN on her first move and OUT on her second move when player 1 plays NOT KNOW. A selfish player 2 plays OUT on both moves when player 1 plays KNOW.

(b) $(\frac{2-9\sigma}{6-9\sigma})^2 < \gamma < \frac{2-9\sigma}{6-9\sigma}$: Player 1 plays KNOW if $\sigma < -\frac{2}{9}$. If $\sigma > -\frac{2}{9}$, player 1 plays KNOW whenever $\gamma < \frac{4-36\sigma+81\sigma^2}{20-36\sigma+81\sigma^2}$ and NOT KNOW whenever $\gamma > \frac{4-36\sigma+81\sigma^2}{20-36\sigma+81\sigma^2}$. Player 1 plays STAY on his first move and plays STAY with probability $\frac{3}{7}$ and STOP with probability $\frac{4}{7}$ on his second move when he plays NOT KNOW. Beliefs in the

first and second information set are $\mu_1 = \gamma$ and $\mu_2 = \gamma$, respectively. When player 1 plays KNOW, he plays STAY on both moves if player 2 is a trustworthy type and STOP on both moves if player 2 is a selfish type. A trustworthy player 2 plays IN on both moves when player 1 plays KNOW or NOT KNOW. A selfish player 2 plays IN with probability $\frac{4\gamma}{(1-\gamma)(2-9\sigma)}$ and OUT with probability $1p - \frac{4\gamma}{(1-\gamma)(2-9\sigma)}$ on her first move and plays OUT on her last move when player 1 plays NOT KNOW. A selfish player 2 plays OUT on both moves when player 1 plays KNOW.

(c) $\gamma < (\frac{2-9\sigma}{6-9\sigma})^2$: Player 1 plays KNOW. Player 1 plays STOP on his first move and plays STAY with probability $\frac{3}{7}$ and STOP with probability $\frac{4}{7}$ on his second move when he plays NOT KNOW. Beliefs in the first and second information set are $\mu_1 = \gamma$ and $\mu_2 = \gamma$, respectively. When player 1 plays KNOW, he plays STAY on both moves if player 2 is a trustworthy type and STOP on both moves if player 2 is a selfish type. A trustworthy player 2 plays IN on both moves when player 1 chooses KNOW or NOT KNOW. A selfish player 2 plays IN with probability $\frac{4\gamma}{(1-\gamma)(2-9\sigma)}$ and OUT with probability $1 - \frac{4\gamma}{(1-\gamma)(2-9\sigma)}$ on her first move and plays OUT on her second move when player 1 plays NOT KNOW. A selfish player 2 plays OUT on both moves when player 1 chooses KNOW.

We now calculate the optimal choice of player 1 in period 2, when $\sigma > \frac{2}{9}$. When player 1 chooses NOT KNOW, we are in the Imperfect game and when he chooses KNOW we are in the Choice game. We have shown that, in the Imperfect game: a trustworthy player 2 will choose IN on both moves and player 1 chooses STAY on both moves in equilibrium, a selfish player 2 chooses IN on his first move, OUT on his second move and player 1 chooses STAY on both moves. If player 1 chooses KNOW in period 2, we have shown that: a trustworthy player 2 will choose IN on both moves and player 1 chooses STAY on both moves in equilibrium, a selfish player 2 chooses IN on his first move, OUT on his second move and player 1 chooses STAY on both moves. This, the equilibrium predictions are the same for the trust game when player 1 chooses KNOW or NOT KNOW on his first move. So that the expected payoff of player 1 is the same when he chooses KNOW or NOT KNOW, implying that he is indifferent between choosing KNOW or NOT KNOW in equilibrium.

Prediction Choice, $\sigma > \frac{2}{9}$. *Let player 2 be a trustworthy type with probability γ and a selfish type with probability $1 - \gamma$. The theoretical predictions of the Choice game are: Player 1 is indifferent between playing KNOW or NOT KNOW. He plays STAY on his both moves. A trustworthy player 2 plays IN on his both moves and a selfish player 2 plays IN on his first move and OUT on his second move.*

4.B. Instructions

General Instructions

You are participating in an experiment on economic decision making and will be asked to make a number of decisions. Please do not talk or communicate in any other way with other participants. If you have a question, raise your hand. Please read these instructions carefully as they describe how you can earn money.

The experiment is anonymous: that is, your identity will not be revealed to others and the identity of others will not be revealed to you.

The experiment consists of 3 parts. At the beginning of each part you will receive instructions describing the game that will be played. All the interaction between you and other participants will take place through the computers.

The experiment is anonymous: that is, your identity will not be revealed to others and the identity of others will not be revealed to you.

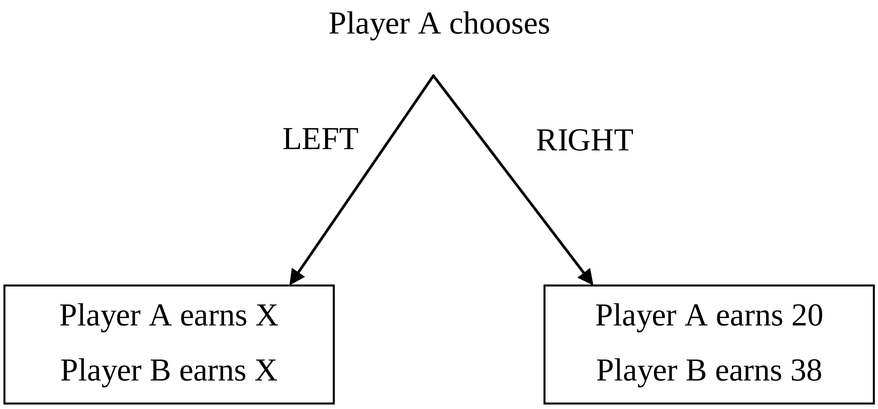
During the experiment your earnings will be expressed in points. Points will be converted to Euros at the following rate: 10 points = 2 EUR. Your total earnings will be the sum of your earnings in each part. You will be paid your earnings in cash at the end of the experiment.

SEE NEXT PAGE

Instructions Part 1

In Part 1 of the experiment, you will be randomly paired with another participant. You will be assigned the role of **Player A** or the role of **Player B**. Each Player A is matched with a Player B. You will get to see which role you are assigned on the computer screen before Part 1 starts.

Consider the decision situation shown in this figure:



If Player A chooses LEFT, both players receive an amount X. If Player A chooses RIGHT, Player A earns 20 points and Player B earns 38 points. The earnings of both players only depend on the decision of Player A. Player B does not make any decision.

The task for **Player A** is to choose a minimum X above which he/she would choose LEFT, where the minimum X should be between 1 and 34. At the end of the experiment, a number X between 1 and 34 will be randomly drawn, and the earnings Player A and Player B receive, depend on this number, and the corresponding choice by Player A of the minimum X.

<u>Example 1</u>: Suppose that Player A chooses a minimum X of 10, and suppose the randomly drawn number X is equal to 20. In this example the minimum X is smaller than the randomly drawn number (10 < 20), which means that Player A prefers to choose LEFT. The earnings will thus be as follows: 20 points for Player A, and 20 points for Player B.

<u>Example 2</u>: Suppose that Player A chooses a minimum X of 20, and suppose the randomly drawn number X is equal to 15. In this example the minimum X is larger than the randomly drawn number (20 > 15), which means that Player A prefers to choose RIGHT. The earnings will thus be as follows: 20 points for Player A, and 38 points for Player B.

Instructions Part 2

In Part 2 of the experiment, you will be randomly paired with another participant. You will be assigned the role of **Player A,** and the matched participant will be assigned the role of Player B.

The game

Your task will be to make decisions for the decision situation presented in the attached figure. There are at most 4 periods. In period 1 Player A chooses either STAY or STOP. If Player A chooses STOP, the game ends, and if Player A chooses STAY, the game continues. In period 2 Player B chooses either IN or OUT. If Player B chooses OUT, the game ends, and if Player B chooses IN, the game continues. In period 3 Player A chooses either STAY or STOP. If Player A chooses STOP, the game ends, and if Player A chooses STAY, the game continues. In period 4 Player B chooses either IN or OUT, and the game ends. The earnings in points of both players, depending on the choices made, are given in the boxes.

As you can see from the figure, Player A has three possible strategies:
(1) STOP in period 1,
or (2) STAY in period 1 and STOP in period 3,
or (3) STAY in period 1 and STAY in period 3.

Also Player B has three possible strategies:
(1) OUT in period 2,
or (2) IN in period 2 and OUT in period 4,
or (3) IN in period 2 and IN in period 4.

This is the sequence in which choices will be asked for:

Step 1: Player B will be asked to choose between the following two options:

- IN in period 2 and IN in period 4

  OR

- OUT in period 2 *or* IN in period 2 and OUT in period 4

Step 2:

- The percentage of participants in the role of Player B who choose IN in period 2 and IN in period 4 in Step 1 in this session will be communicated to all participants (Player A and Player B)

- Player A (you) will get the choice to get information about the choice by the matched Player B made in Step 1. You will be asked: Do you want to know which choice Player B made in Step 1?

  If you answer Yes to this question, you get information about B's decision in Step 1.

  If you answer No to this question, you get no information about B's decision in Step 1.

  Player B will be informed about this communication at the point he/she is asked to make a choice in Step 3, ***only if you answer Yes to this question***.

Step 3:

Choice Player A: Player A will be asked to choose between the following three options:

- STAY in period 1 and STAY in period 3

  OR

- STAY in period 1 and STOP in period 3

  OR

- STOP in period 1

Choice Player B:

- If Player B has chosen IN in period 2 and IN in period 4 (in Step 1), then there is no choice.
- If Player B has ***not*** chosen IN in period 2 and IN in period 4 (in Step 1), then Player B will be asked to choose between the following two options:
  - IN in period 2 and OUT in period 4

    OR

  - OUT in period 2

Number of rounds

The task described will be repeated 10 times. At the beginning of each of the 10 rounds, the pairs will be randomly reshuffled.

In each round, choices will be asked as described above. However, the choice made by Player B in Step 1 will be kept the same in all of the 10 rounds.

At the end of the experiment, one out of the 10 rounds will be randomly drawn for payment.

Example

Step 1: Suppose that Player B chooses IN in period 2 and IN in period 4.

Step 2: Suppose that 3 out of 16 participants in the role of Player B have chosen IN in period 2 and IN in period 4 (in Step 1). This information will be shown on the computer screen. Additionally, if Player

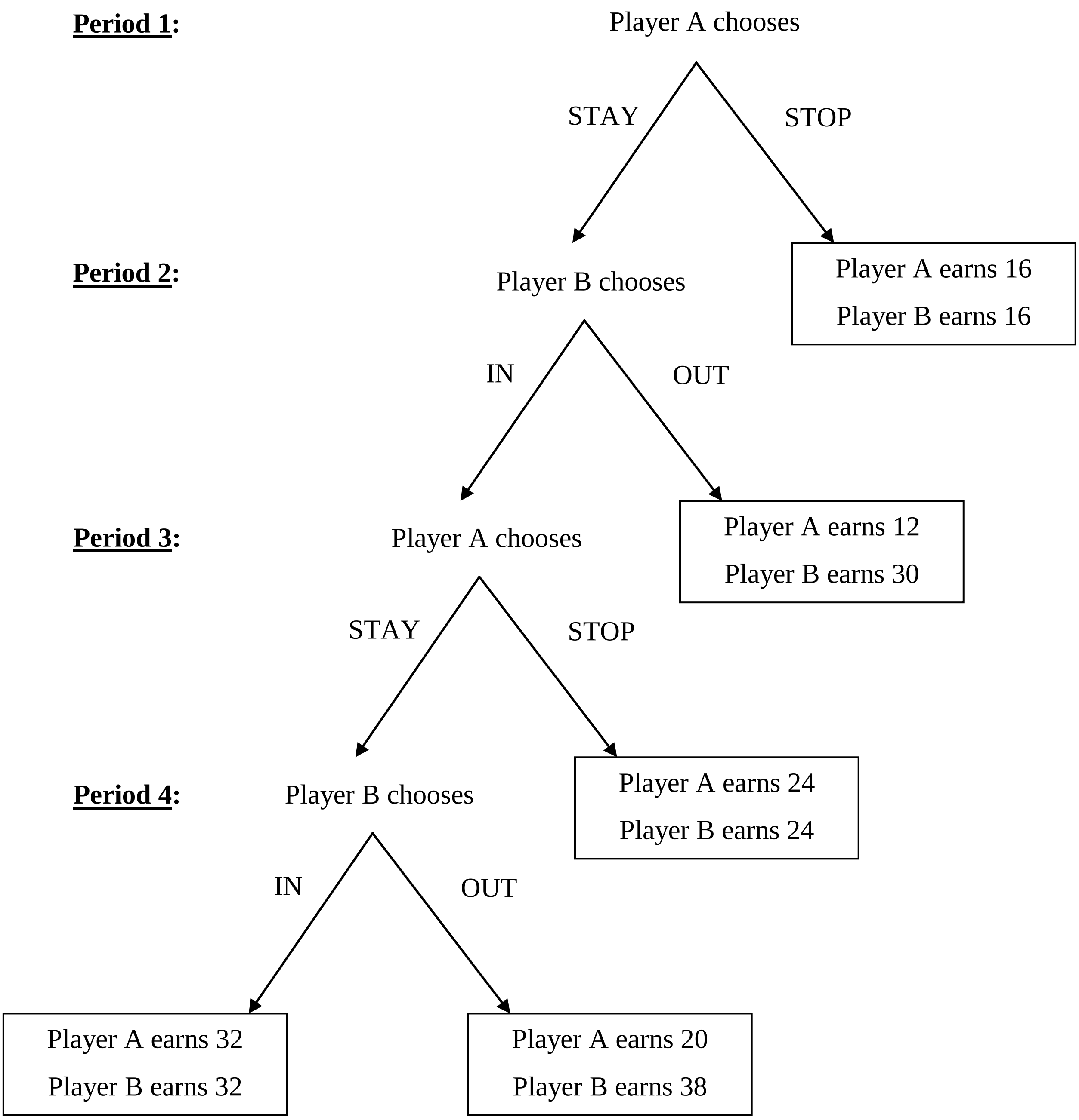
A chooses to get information about the matched Player B's choice in Step 1, this information is also shown.

Step 3: Suppose that Player A chooses STAY in period 1 and STAY in period 3.

In this example, earnings would be calculated as follows:

- Period 1: Player A chooses STAY. Therefore, the game continues to period 2.
- Period 2: Player B chooses IN. Therefore, the game continues to period 3.
- Period 3: Player A chooses STAY. Therefore, the game continues to period 4.
- Period 4: Player A chooses IN.

In this example, Player A gets 32 points and Player B gets 32 points.



Instructions Part 2

In Part 2 of the experiment, you will be randomly paired with another participant. You will be assigned the role of **Player B,** and the matched participant will be assigned the role of Player A.

The game

Your task will be to make decisions for the decision situation presented in the attached figure. There are at most 4 periods. In period 1 Player A chooses either STAY or STOP. If Player A chooses STOP, the game ends, and if Player A chooses STAY, the game continues. In period 2 Player B chooses either IN or OUT. If Player B chooses OUT, the game ends, and if Player B chooses IN, the game continues. In period 3 Player A chooses either STAY or STOP. If Player A chooses STOP, the game ends, and if Player A chooses STAY, the game continues. In period 4 Player B chooses either IN or OUT, and the game ends. The earnings in points of both players, depending on the choices made, are given in the boxes.

As you can see from the figure, Player A has three possible strategies:
(1) STOP in period 1,
or (2) STAY in period 1 and STOP in period 3,
or (3) STAY in period 1 and STAY in period 3.

Also Player B has three possible strategies:
(1) OUT in period 2,
or (2) IN in period 2 and OUT in period 4,
or (3) IN in period 2 and IN in period 4.

This is the sequence in which choices will be asked for:

Step 1: Player B will be asked to choose between the following two options:

- IN in period 2 and IN in period 4

  OR

- OUT in period 2 *or* IN in period 2 and OUT in period 4

Step 2: The percentage of participants in the role of Player B who choose IN in period 2 and IN in period 4 in Step 1 in this session will be communicated to all participants (Player A and Player B).

<u>Step 3</u>:

<u>Choice Player A</u>: Player A will be asked to choose between the following three options:

- STAY in period 1 and STAY in period 3

  OR

- STAY in period 1 and STOP in period 3

  OR

- STOP in period 1

<u>Choice Player B</u>:

- If Player B has chosen IN in period 2 and IN in period 4 (in Step 1), then there is no choice.
- If Player B has ***not*** chosen IN in period 2 and IN in period 4 (in Step 1), then Player B will be asked to choose between the following two options:
  - IN in period 2 and OUT in period 4

    OR

  - OUT in period 2

Number of rounds

The task described will be repeated 10 times. At the beginning of each of the 10 rounds, the pairs will be randomly reshuffled.

In each round, choices will be asked as described above. However, the choice made by Player B in Step 1 will be kept the same in all of the 10 rounds.

At the end of the experiment, one out of the 10 rounds will be randomly drawn for payment.

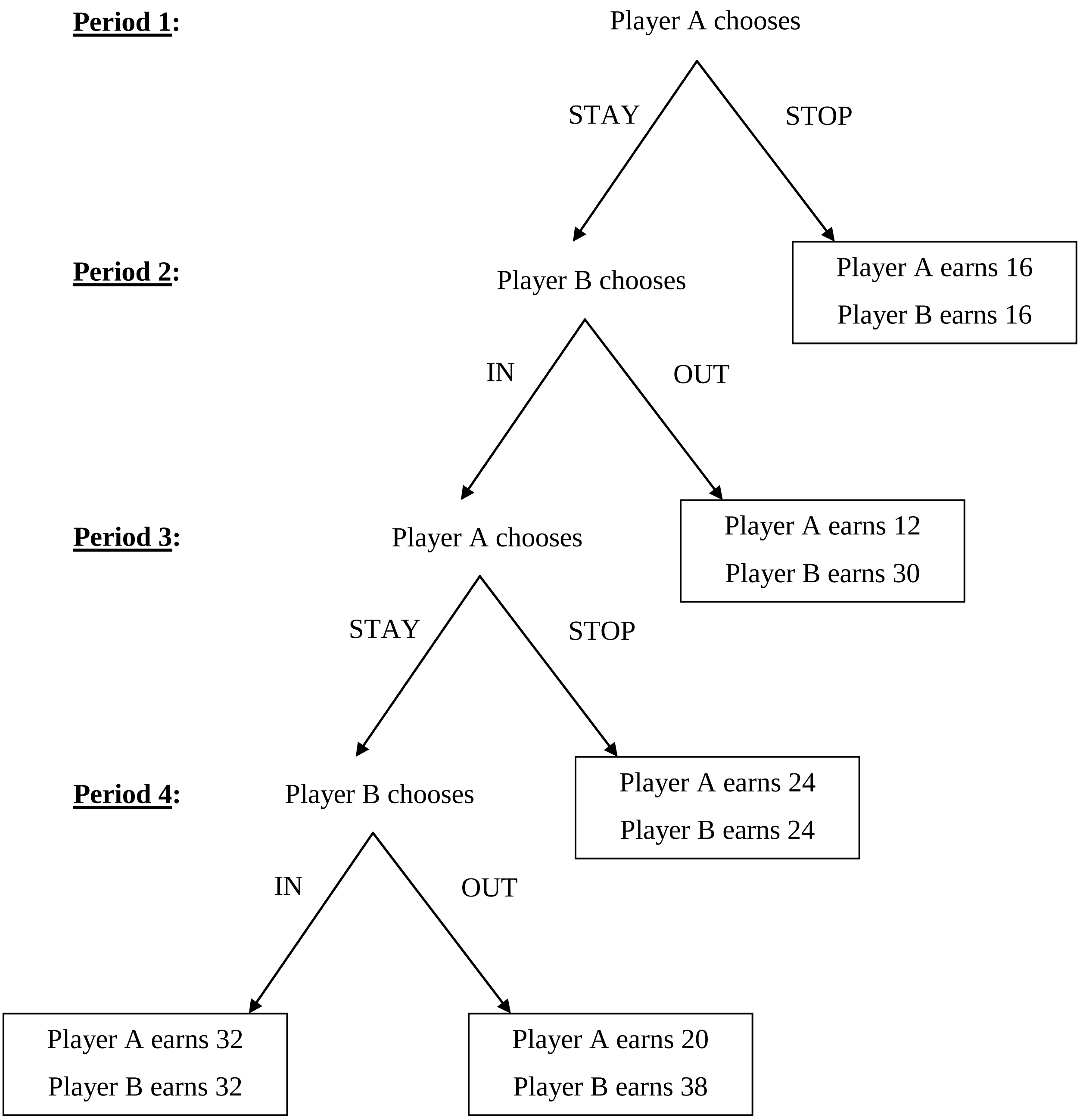
Example

<u>Step 1</u>: Suppose that Player B chooses IN in period 2 and IN in period 4.
<u>Step 2</u>: Suppose that 3 out of 16 participants in the role of Player B have chosen IN in period 2 and IN in period 4 (in Step 1). This information will be shown on the computer screen.
<u>Step 3</u>: Suppose that Player A chooses STAY in period 1 and STAY in period 3.

In this example, earnings would be calculated as follows:

- Period 1: Player A chooses STAY. Therefore, the game continues to period 2.
- Period 2: Player B chooses IN. Therefore, the game continues to period 3.
- Period 3: Player A chooses STAY. Therefore, the game continues to period 4.
- Period 4: Player A chooses IN.

In this example, Player A gets 32 points and Player B gets 32 points.



Instructions Part 2

In Part 2 of the experiment, you will be randomly paired with another participant. You will be assigned the role of **Player A** or **Player B,** and the matched participant will be assigned the other role.

The game

Your task will be to make decisions for the decision situation presented in the attached figure. There are at most 4 periods. In period 1 Player A chooses either STAY or STOP. If Player A chooses STOP, the game ends, and if Player A chooses STAY, the game continues. In period 2 Player B chooses either IN or OUT. If Player B chooses OUT, the game ends, and if Player B chooses IN, the game continues. In period 3 Player A chooses either STAY or STOP. If Player A chooses STOP, the game ends, and if Player A chooses STAY, the game continues. In period 4 Player B chooses either IN or OUT, and the game ends. The earnings in points of both players, depending on the choices made, are given in the boxes.

As you can see from the figure, Player A has three possible strategies:
(1) STOP in period 1,
or (2) STAY in period 1 and STOP in period 3,
or (3) STAY in period 1 and STAY in period 3.

Also Player B has three possible strategies:
(1) OUT in period 2,
or (2) IN in period 2 and OUT in period 4,
or (3) IN in period 2 and IN in period 4.

This is the sequence in which choices will be asked for:

Step 1: Player B will be asked to choose between the following two options:

- IN in period 2 and IN in period 4

  OR

- OUT in period 2 *or* IN in period 2 and OUT in period 4

Step 2: The percentage of participants in the role of Player B who choose IN in period 2 and IN in period 4 in Step 1 in this session will be communicated to all participants (Player A and Player B).

Step 3:

Choice Player A: Player A will be asked to choose between the following three options:

- STAY in period 1 and STAY in period 3

  OR

- STAY in period 1 and STOP in period 3

  OR

- STOP in period 1

Choice Player B:

- If Player B has chosen IN in period 2 and IN in period 4 (in Step 1), then there is no choice.
- If Player B has ***not*** chosen IN in period 2 and IN in period 4 (in Step 1), then Player B will be asked to choose between the following two options:
  - IN in period 2 and OUT in period 4

    OR

  - OUT in period 2

Number of rounds

The task described will be repeated 10 times. At the beginning of each of the 10 rounds, the pairs will be randomly reshuffled.

In each round, choices will be asked as described above. However, the choice made by Player B in Step 1 will be kept the same in all of the 10 rounds.

At the end of the experiment, one out of the 10 rounds will be randomly drawn for payment.

Example

Step 1: Suppose that Player B chooses IN in period 2 and IN in period 4.

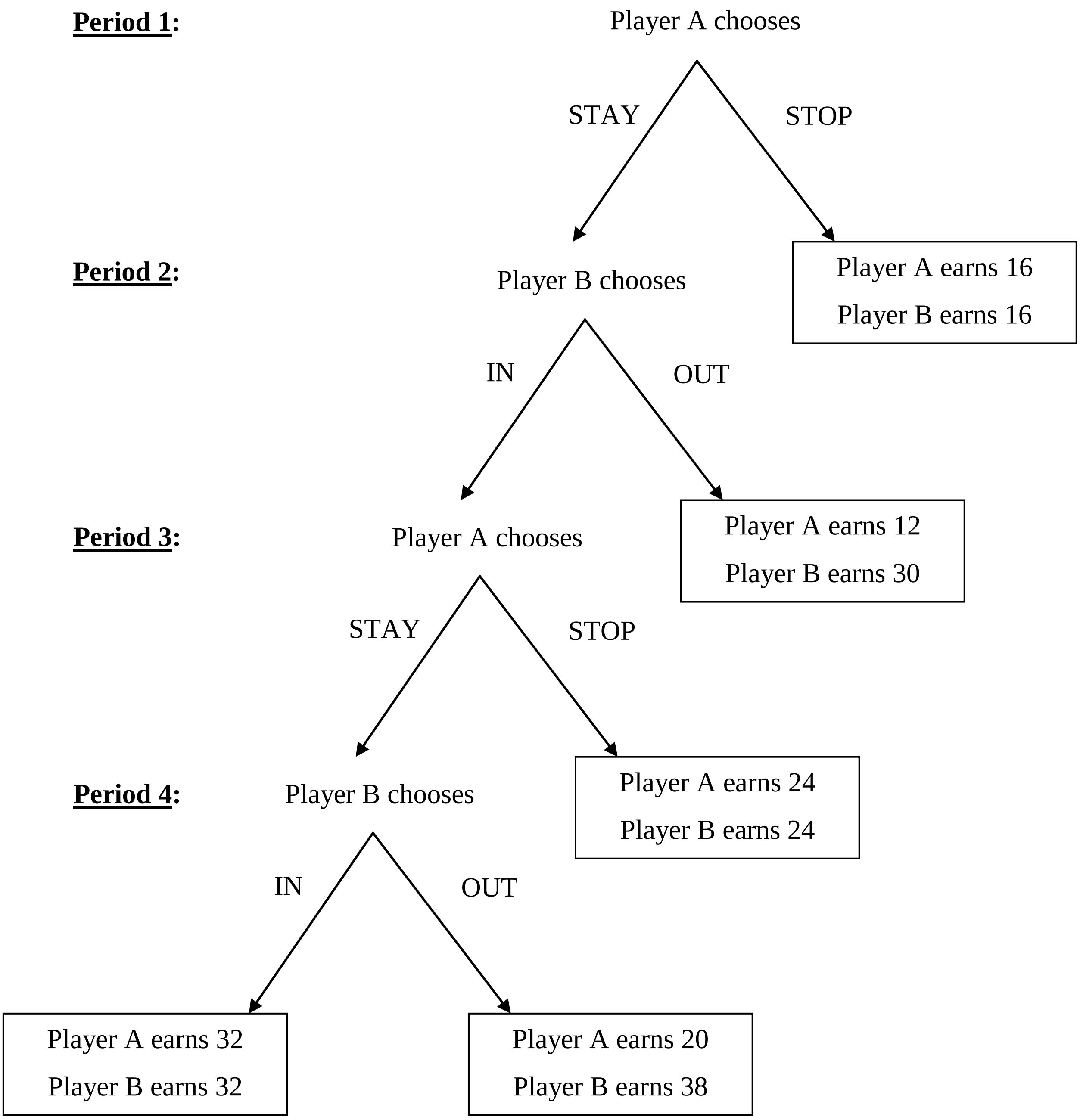
Step 2: Suppose that 3 out of 16 participants in the role of Player B have chosen IN in period 2 and IN in period 4 (in Step 1). This information will be shown on the computer screen.

Step 3: Suppose that Player A chooses STAY in period 1 and STAY in period 3.

In this example, earnings would be calculated as follows:

- Period 1: Player A chooses STAY. Therefore, the game continues to period 2.
- Period 2: Player B chooses IN. Therefore, the game continues to period 3.
- Period 3: Player A chooses STAY. Therefore, the game continues to period 4.
- Period 4: Player A chooses IN.

In this example, Player A gets 32 points and Player B gets 32 points.



Instructions Part 3

In Part 3 of the experiment, the same decision situation as in Part 2 appears, and all participants keep the same role as in Part 2.

The same task as the one in Part 2 will be repeated another 10 times. In each round, choices will be asked as described in Part 2. However, Player B has no choice to make in Step 1 because this choice is kept the same as in Part 2 of the experiment in all of the 10 rounds. At the beginning of each of the 10 rounds, the pairs will be randomly reshuffled.

Now, in Step 2, Player A (you) will get the choice to get information about the choice by the matched Player B made in Step 1. You will be asked: Do you want to know which choice Player B made in Step 1?

- If you answer Yes to this question, you get information about B's decision in Step 1.
- If you answer No to this question, you get no information about B's decision in Step 1.

Player B will be informed about this communication at the point he/she is asked to make a choice in Step 3, ***only if you answer Yes to this question***.

At the end of the experiment, one out of the 10 rounds will be randomly drawn for payment.

Instructions Part 3

In Part 3 of the experiment, the same decision situation as in Part 2 appears, and all participants keep the same role as in Part 2.

The same task as the one in Part 2 will be repeated another 10 times. In each round, choices will be asked as described in Part 2. However, Player B (you) has no choice to make in Step 1 because this choice is kept the same as in Part 2 of the experiment in all of the 10 rounds. At the beginning of each of the 10 rounds, the pairs will be randomly reshuffled.

At the end of the experiment, one out of the 10 rounds will be randomly drawn for payment.

Instructions Part 3

In Part 3 of the experiment, the same decision situation as in Part 2 appears, and all participants keep the same role as in Part 2.

The same task as the one in Part 2 will be repeated another 10 times. In each round, choices will be asked as described in Part 2. However, Player B has no choice to make in Step 1 because this choice is kept the same as in Part 2 of the experiment in all of the 10 rounds. At the beginning of each of the 10 rounds, the pairs will be randomly reshuffled.

Another difference is that now, in Step 2, Player A will no longer obtain information about the choice by the matched Player B made in Step 1, and Player B knows this. (The instructions are the same for everyone.)

At the end of the experiment, one out of the 10 rounds will be randomly drawn for payment.

Bibliography

Abeler, J., A. Falk, L. Goette, and D. Huffman (2011). Reference points and effort provision. *The American Economic Review 101*(2), 470–492.

Aimone, J. A. and D. Houser (2012). What you don't know won't hurt you: a laboratory analysis of betrayal aversion. *Experimental Economics 15*, 571–588.

Aimone, J. A. and D. Houser (2013). Harnessing the benefits of betrayal aversion. *Journal of Economic Behavior & Organization 89*, 1–8.

Anderhub, V., D. Engelmann, and W. Güth (2002). An experimental study of the repeated trust game with incomplete information. *Journal of Economic Behavior and Organization 48*, 197–216.

Barut, Y., K. D. and C. Noussair (2002). A comparison of multiple unit all pay auction and winner pay auction under incomplete information. *International Economic Review 43*(3), 675–708.

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

Blonski, M., P. Ockenfels, and G. Spagnolo (2011). Equilibrium selection in the repeated prisoner's dilemma: Axiomatic approach and experimental evidence. *American Economic Journal: Microeconomics 3*, 164–192.

Blonski, M. and G. Spagnolo (in press). Prisoners' other dilemma. *International Journal of Game Theory*.

Bohnet, I. and R. Zeckhauser (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization* (4), 467–484.

Bolton, G. E. and A. Ockenfels (2000). Erc: A theory of equity, reciprocity, and competition. *American Economic Review*, 166–193.

Brandts, J. and N. Figueras (2003). An exploration of reputation formation in experimental games. *Journal of Economic Behavior & Organization 50*, 89–115.

Bulow, J. I., J. Genakoplos, and D. Klemperer (1985). Multimarket oligopoly: Strategic substitutes and complements. *Journal of Political Economy 93*, 488–511.

Camerer, C. and E. Fehr (2006). When does “economic man” dominate social behavior? *Science 711*, 47–52.

Camerer, C. and K. Weigelt (1988). Experimental tests of a sequential equilibrium reputation model. *Econometrica 56*, 1–36.

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

Clark, K. and M. Sefton (2001). The sequential prisoner’s dilemma: Evidence on reciprocation. *Economic Journal 111*, 51–68.

Crawford, V. P. and J. Meng (2011). New york city cab drivers’ labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income. *The American Economic Review 101*(5), 1912–1932.

Dal Bó, P. and G. R. Fréchette (2011). The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review 101*, 411–429.

Delgado, M. R., A. Schotter, E. Y. Ozbay, and E. A. Phelps (2008). Understanding overbidding: Using the neural circuitry of reward to design economic auctions. *Science 321*(5897), 1849–1852.

Dominguez-Martinez, S., R. Sloof, and F. A. von Siemens (2014). Monitored by your friends, not your foes: Strategic ignorance and the delegation of real authority. *Games and Economic Behavior*, 289–305.

Duffy, J. and F. Munoz-García (2013). Cooperation and signaling with uncertain social preferences. *Theory and Decision*, Forthcoming.

Eisenhuth, R. and M. Ewers (2012). Auction with loss averse bidders. working paper.

Embrey, M., F. Mengel, and R. Peeters (2014). Strategic commitment and cooperation in experimental games of strategic complements and substitutes. Working Paper.

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

Ernst, C. and C. Thöni (2009). *Bimodal Bidding in Experimental All-Pay Auctions*. Department of Economics, Univ.

Fehr, E. and U. Fischbacher (2002). Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives. *Economic Journal 112*, C1–C33.

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

Fehr, E. and J. Tyran (2008). Limited rationality and strategic interaction. the impact of the strategic environment on nominal inertia. *Econometrica 76*, 353–394.

Fischbacher, U. (2007a). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*, 171–178.

Fischbacher, U. (2007b). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*, 171–178.

Freeman, R. and A. Gelber (2009). Prize structure and information in tournaments: Experimental evidence. *American Economic Journal: Applied Economics 2*(1), 149–164.

Friedman, J. W. (1971). A non-cooperative equilibrium for supergames. *Review of Economic Studies 38*, 1–12.

Fudenberg, D. and J. Tirole (1984). The fat-cat effect, the puppy-dog ploy, and the lean and hungry look. *American Economic Review 74*, 361–366.

Gazzale, R. S. (2009). Learning to play nash from the best. Technical report.

Guiso, L., P. Sapienza, and L. Zingales (2004). The role of social capital in financial development. *American Economic Review 94*, 526–556.

Guiso, L., P. Sapienza, and L. Zingales (2008). Trusting the stock market. *The Journal of Finance 63*, 2557–2600.

Guiso, L., P. Sapienza, and L. Zingales (2009). Cultural biases in economic exchange? *The Quarterly Journal of Economics 124*(3), 1095–1131.

Haltiwanger, J. and M. Waldman (1985). Rational expectations and the limits of rationality: An analysis of heterogeneity. *American Economic Review 75*, 326–340.

Haltiwanger, J. and M. Waldman (1991). Responders versus non-responders: A new perspective on heterogeneity. *Economic Journal 101*, 1085–1102.

Haltiwanger, J. and M. Waldman (1993). The role of altruism in economic interaction. *Journal of Economic Behavior & Organization 21*, 1–15.

Herweg, F., D. Müller, and P. Weischenk (2010). Binary payment schemes: Moral hazard and loss aversion. *American Economic Review 100*(5), 2451–2477.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–291.

Kandori, M. (1992). Social norms and community enforcement. *Review of Economic Studies 59*, 63–80.

Kőszegi, B. (2013). Behavioral contract theory? working paper.

Kőszegi, B. and M. Rabin (2006). A model of reference-dependent preferences. *Quarterly Journal of Economics 121*(4), 1133–1165.

Klose, B. and R. Sheremeta (2012). Behavior in all-pay and winner-pay auctions with identity-dependent externalities.

Kreps, D. M., P. Milgrom, J. Roberts, and R. Wilson (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory 27*, 245–252.

Lange, A. and A. Ratan (2010). Multi-dimensional reference-dependent preferences in sealed bid auctions - how (most) laboratory experiments differ from the field. *Games and Economic Behavior 68*(2), 634–645.

McKelvey, R. D. and T. R. Palfrey (1992). An experimental study of the centipede game. *Econometrica 60*, 803–836.

Moldovanu, B. and A. Sela (2001). The optimal allocation of prizes in contests. *American Economic Review 91*(3), 542–558.

Müller, W. and A. Schotter (2010). Workaholics and dropouts in organizations. *Journal of the European Economic Association 8*(4), 717–743.

Neral, J. and J. Ochs (1992). The sequential equilibrium theory of reputation building: A further test. *Econometrica 60*, 1151–1169.

Noussair, C. and J. Silver (2006). Behavior in all-pay auctions with incomplete information. *Games and Economic Behavior 55*(1), 189–206.

Post, T., M. J. Van den Assem, G. Baltussen, and R. H. Thaler (2008). Deal or no deal? decision making under risk in a large-payoff game show. *The American economic review*, 38–71.

Potters, J. and S. Suetens (2009). Cooperation in experimental games of strategic complements and substitutes. *Review of Economic Studies 76*, 1125–1147.

Reuben, E. and S. Suetens (2009). On the cognitive and motivational components of reciprocity. Working Paper.

Reuben, E. and S. Suetens (2012). Revisiting strategic versus non-strategic cooperation. *Experimental Economics 15*, 24–43.

Reuben, E. and S. Suetens (2014a). Instrumental reciprocity as an error.

Reuben, E. and S. Suetens (2014b). Signaling trustworthiness. mimeo.

Roth, A. E. and J. K. Murnighan (1978). Equilibrium behavior and repeated play of the prisoner's dilemma. *Journal of Mathematical Psychology 17*, 189–198.

Schmidt, D., R. Shupp, J. Walker, and E. Ostrom (2003). Playing safe in coordination games: The roles of risk dominance, payoff dominance, and history of play. *Games and Economic Behavior 42*, 281–299.

Schram, A. J. H. C. and S. Onderstal (2009). Bidding to give: An experimental comparison of auctions for charity. *International Economic Review 50*, 431–457.

Topkis, D. M. (1978). Minimizing a submodular function on a lattice. *Operations Research 26*, 305–321.

Van Huyck, J. B., R. C. Battalio, and R. O. Beil (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review 80*, 234–248.

Zak, P. J. and S. Knack (2001). Trust and growth. *The Economic Journal 111*, 295–321.