Experiments in Competition, Cooperation, and Learning

vorgelegt von
Mag.rer.soc.oec.
Johannes Josef Leutgeb
geb. in Linz (Österreich)

von der Fakultät VII - Wirtschaft und Management
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften
- Dr. rer. oec. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzende:	Prof. Dr. Radosveta Ivanova-Stenzel
Gutachterin:	Prof. Dr. Dorothea Kübler
Gutachter:	Prof. Dr. Steffen Huck

Tag der wissenschaftlichen Aussprache: 5. Dezember 2017

Berlin 2018

Angaben zu bereits veröffentlichten Teilen der Dissertation

Diese Dissertation besteht aus drei Kapiteln. Kapitel 1 ist die Postprint-Version eines Artikels, der bereits in der Zeitschrift Nature Communications publiziert wurde. Der Verlag Nature Publishing Group erlaubt die Verwendung der Postprint-Version für Dissertationen.

Huck, Steffen, Johannes Leutgeb und Ryan Oprea (2017): Payoff information hampers the evolution of cooperation. Nature Communications 8 (15147). https://doi.org/10.1038/ncomms15147

Kapitel 2 ist bisher unveröffentlicht.

Kapitel 3 ist die Preprint-Version eines Artikels, der bereits als Discussion Paper in der Reihe der WZB Discussion Papers veröffentlicht wurde. Am 16. März 2018 wurde der Artikel bei der Zeitschrift Games and Economic Behavior (Elsevier Verlag) eingereicht.

Ensthaler, Ludwig, Steffen Huck und Johannes Leutgeb (2017): Games played through agents in the laboratory – A test of Prat & Rustichini's model. WZB Discussion Papers, Research Unit: Economics of Change SP II 2016-305r, Social Science Research Center Berlin (WZB). https://bibliothek.wzb.eu/pdf/2017/ii16-305r.pdf

Statement on previously published parts of the dissertation

This dissertation has three chapters. Chapter 1 is the postprint version of an article which was previously published in the journal Nature Communications. The publisher Nature Publishing Group allows the use of the postprint version for theses.

Huck, Steffen, Johannes Leutgeb and Ryan Oprea (2017): Payoff information hampers the evolution of cooperation. Nature Communications 8 (15147). https://doi.org/10.1038/ncomms15147

Chapter 2 has not yet been published.

Chapter 3 is the preprint version of an article which was previously published as a discussion paper in the WZB Discussion Papers series. On 16 March 2018 the article was submitted to the journal Games and Economic Behavior (publisher Elsevier).

Ensthaler, Ludwig, Steffen Huck and Johannes Leutgeb (2017): Games played through agents in the laboratory – A test of Prat & Rustichini's model. WZB Discussion Papers, Research Unit: Economics of Change SP II 2016-305r, Social Science Research Center Berlin (WZB). https://bibliothek.wzb.eu/pdf/2017/ii16-305r.pdf

Zusammenfassung

Ich untersuche menschliches Lernverhalten in ökonomischen Umgebungen unter Verwendung von ökonomischen Experimenten in einer kontrollierten Laborumgebung. Kapitel 1 und 2 behandeln Lernverhalten in einem opaken Cournotmarkt, in dem Teilnehmer in fixen Gruppen über hunderte Perioden interagieren. Kapitel 3 dreht sich um die Frage, wie Teilnehmer lernen ein komplexes Spiel um Einfluss zu spielen. Teilnehmer folgen einfachen Lernregeln um durch komplizierte und opake ökonomische Umgebungen zu navigieren. Abhängig von der Informationsumgebung können diese Lernregeln Teilnehmer zur Kooperation oder zum Wettbewerb führen.

Abstract

I investigate human learning in economic environments using economic experiments in a controlled laboratory environment. Chapters 1 and 2 investigate learning behavior in an opaque Cournot environment in which subjects interact in fixed groups for hundreds of periods. Chapter 3 deals with the question of how subjects learn to play a complex game of influence. Subjects follow simple learning rules to navigate complicated and opaque economic environments. Depending on the informational environment, these learning rules can lead subjects towards cooperation or competition.

Contents

List of Tables

List of Figures

Preface

Making economic decision is no easy task. Standard economic models typically assume that agents possess full information on how the economic environment in which they interact with others works. Even in standard models of asymmetric information agents are assumed to have (rational) expectations on the possible states of the environment. In addition, agents are usually assumed to possess infinite mental capacity to analyze and understand their environment in order to make an optimal decision. In the last decades scholars in behavioral economics have started to scrutinize these assumptions, leading to the question of how to conceptualize human learning processes. Economists have proposed and employed rational as well as non-rational models of learning. This dissertation contains three chapters that all use methods from experimental economics or analyze experimental data. Experimental methods lend themselves to studying learning because they allow us to disentangle different modes of learning and clearly identify behavioral patterns. All three chapters in this dissertation investigate human learning behavior.

The first chapter is based on work with Steffen Huck and Ryan Oprea. In Huck et al. (2017) we investigate the role of payoff information for the evolution of cooperation. Subjects interact repeatedly on Cournot markets in duopolies with very little information available to them. They only know that the function that generates their payoffs depends on their own and their partner's action, that it will remain stable over time and that it is symmetric. Subjects receive feedback that allows them to learn though. After each interaction they learn of the combination of choices on the market and the resulting payoffs for both participants. Additionally, we introduce a treatment variation and give subjects in some sessions additional payoff information on what they could have earned in the previous period. First, we establish that subjects follow simple learning rules. In the baseline treatments with little information subjects initially start to imitate actions that yield relatively higher payoffs, leading to highly competitive choices above the Nash equilibrium predictions. Over time, subjects manage to abandon this heuristic and employ more cooperative learning rules like win-continue lose-reverse (WCLR) (Huck et al., 2003) that lead them, eventually, to cooperation. Furthermore, we show that such learning rules are sensitive to the informational environment. In treatments with more information subjects are much less likely to cooperate. This chapter establishes how simple learning rules can foster cooperation between humans.

In the second chapter I investigate the learning process towards cooperation more closely and I identify a possibly competing explanation. I argue that a subject's adjustment towards a lower action is a kind act as doing so unambiguously increases her partner's payoff. I introduce a novel idea of reciprocity not in actions but in adjustments of actions and test WCLR against reciprocity in the data. I find that, while in some cases the learning rules predict the same movement, both play a role in subjects' behavior. Then I use this concept of reciprocity to better understand heterogeneity in the data. I separate groups by their characteristics and I find that female groups are less likely to find cooperative agreements than men groups or mixed gender groups. Supporting these findings on the group level, I find women on the individual level to be less likely to reciprocate kind actions than men. I connect these findings to the mixed results literature on gender differences in reciprocity in the literature. These results provide a further piece of evidence on the question of gender differences in social preferences.

Finally, the third chapter is based on joint work with Ludwig Ensthaler and Steffen Huck (Ensthaler et al., 2017). In this chapter, we test a model of influence by Prat and Rustichini (2003) in the laboratory. They formulate an interesting but complex model of principals who compete for the influence over some agents. In equilibrium, the agents implement the efficient outcome, supported by principals' offers that are the result of intricate strategies. We ask if such a complex theory has predictive power in the laboratory and we design a laboratory experiment with human principals and computerized agents. Do subjects get it right, and if not initially, do they learn to play equilibrium? We find that, on the level of outcomes, the theory does remarkably well. The computzerized agents implement the efficient outcome most of the time. The offers by the human principals do not follow the theoretical predictions entirely. First, we find that there is learning over time. Subjects adjust their offers in a process that can be described by learning direction theory (Selten et al., 2005). After receiving feedback about their choices and the resulting outcomes in the previous period, subjects move myopically towards actions that would have resulted in higher payoffs in the previous period. This process doesn't lead subjects towards the Nash equilibrium in pure strategies that the theory predicts. Instead, subjects continue to make errors. We argue that subjects are more likely to make errors that hurt them less and we fit quantal response equilibria (QRE). QRE describes behavior well, which is supported by an additional treatment testing QRE against standard predictions.

To summarize, this dissertation documents the power and importance of learning. Using simple learning rules, humans can navigate complicated economic environments and through gradual adjustment learn to play the underlying game. The experiments show that the end result of these learning processes importantly depends on the knowledge that agents have about their environment and the feedback that they receive after having made a choice. Learning rules may support standard theoretical predictions but they may also lead agents astray. There is much future research to be done on the question under which conditions learning supports standard theory and under which not.

Angabe zum bereits veröffentlichtem Kapitel

Das folgende Kapitel ist die Postprint-Version eines Artikels, der bereits in der Zeitschrift Nature Communications publiziert wurde. Der Verlag Nature Publishing Group erlaubt die Verwendung der Postprint-Version für Dissertationen.

Huck, Steffen, Johannes Leutgeb und Ryan Oprea (2017): Payoff information hampers the evolution of cooperation. Nature Communications 8 (15147). https://doi.org/10.1038/ncomms15147

Statement on previously published chapter

The following chapter is the postprint version of an article which was previously published in the journal Nature Communications. The publisher Nature Publishing Group allows the use of the postprint version for theses.

Huck, Steffen, Johannes Leutgeb and Ryan Oprea (2017): Payoff information hampers the evolution of cooperation. Nature Communications 8 (15147). https://doi.org/10.1038/ncomms15147

1 Payoff Information Hampers the Evolution of Cooperation

Abstract

Human cooperation has been explained through rationality- as well as heuristics-based models. Both model classes share the feature that knowledge of payoff functions is weakly beneficial for the emergence of cooperation. Here, we present experimental evidence to the contrary. We let human subjects interact in a competitive environment and find that, in the long run, access to information about own payoffs leads to less cooperative behaviour. In the short run subjects use naïve learning heuristics that get replaced by better adapted heuristics in the long run. With more payoff information subjects are less likely to switch to pro-cooperative heuristics. The results call for the development of two-tier models for the evolution of cooperation.

1.1 Introduction

Grasping the forces and conditions that foster cooperative behaviour is of central importance for the understanding of how societies thrive or decline. Cooperation has been explained through models with purely selfish agents who exhibit high levels of sophistication and reasoning in repeated-game environments (Friedman, 1971; Dal Bó and Fréchette, 2011); static models invoking social preferences, altruistic punishments and reciprocity (Fehr and Schmidt, 1999; Fehr and Gächter, 2002; Gintis et al., 2003; Fehr et al., 2002; Bowles and Gintis, 2002; Dufwenberg and Kirchsteiger, 2004; Wedekind and Milinski, 2000; Rockenbach and Milinski, 2006; Falk and Fischbacher, 2006); evolutionary models that show under which conditions cooperative behaviour, reciprocity and pro-social preferences survive (Trivers, 1971; Axelrod and

Hamilton, 1981; Nowak and Sigmund, 1998; Huck and Oechssler, 1999; Nowak, 2006); and models showing how learning rules and heuristics such as win-stay, lose-shift or win-continue, lose-reverse can generate cooperative behaviour (Nowak and Sigmund, 1993; Oechssler, 2002; Huck et al., 2003, 2004). All these models either depend on the assumption that agents have full knowledge of their own payoff function (i.e., that they know how all possible outcomes in the game they play map into success), or, where such knowledge is not required, imply that having access to payoff information cannot hurt. Here we present evidence to the contrary: we show that in long-run learning environments payoff information hampers the evolution of cooperation. Our evidence suggests that, for understanding behaviour in the long run, two-tier models combining elements of evolutionary models and models of learning rules and heuristics are appropriate. In the short run we find that subjects' behaviour is best described by naïve learning heuristics but in the long-run these heuristics are replaced by better adapted heuristics. We show that subjects are much less likely to learn to adopt heuristics that foster cooperation if they have access to payoff information.

1.2 Results

1.2.1 The Experiment

We study (learning) behaviour in a game with 600 periods that each last eight seconds. (We also have data on games with 1,200 four-second and 2,400 two-second periods which generate similar results. See figures 17 and 18) as well as the supplementary analysis in appendix A.) The game is a two-player contest game with a continuous action space that exhibits a strong tension between competition and cooperation. The payoff function is $\pi_i(x_i, x_{-i}) = 10 + (\frac{120}{\sum_j x_j} - 10)x_i$ where x_i denotes the action of agent i; maximum competition implies $x_i = 6$ with profits of 0.5 euro cents per period, maximum cooperation $x_i = 0.1$ with profits of 3.45 cents per period, and Nash in between at $x_i = 3$ and profits of 2 cents per period. Points are converted into Euro by dividing them by 2000. The one-period Nash equilibrium outcome is in the middle of the action space with an effort of 3 and profits of 2 cents. There are two treatments, Info and NoInfo, varying the feedback information that subjects receive between rounds. In both treatments, subjects observe the action profile of the last period and the resulting payoffs for themselves and their partner. In treatment Info they additionally observe a curve showing the possible payoffs they would have earned in the previous period for alternative choices. Note that the NoInfo treatment is based on previous work by Friedman et al. (2015) that use the same software package albeit slightly different parameters with 4 seconds per period and rematching of subjects every 400 periods. Our findings successfully replicate their findings.

Running this many periods of interaction in a single laboratory session is made feasible using a graphical software interface (see figure 1) where choices are implemented through a slider. Actions are determined by the slider's position at the end of the timed period and the slider remains in the same position when the next period starts. Subjects are shown a progress bar filling up every eight seconds and a counter of remaining periods. In the first period sliders are set at a random position which determines play in that period. After a period has ended subjects receive feedback on both players' actions and payoffs in the previous period: two dots in the interface indicate (through their horizontal position) actions and (through their vertical position) payoffs. Furthermore in treatment Info subjects are shown a curve of potential payoffs which they could have earned given their partner's choice. Notice that subjects are, of course, free to leave the slider in the same position for multiple rounds, in which case actions from previous periods are simply repeated. Indeed, many subjects often choose to wait a little between adjustments which renders the game much less

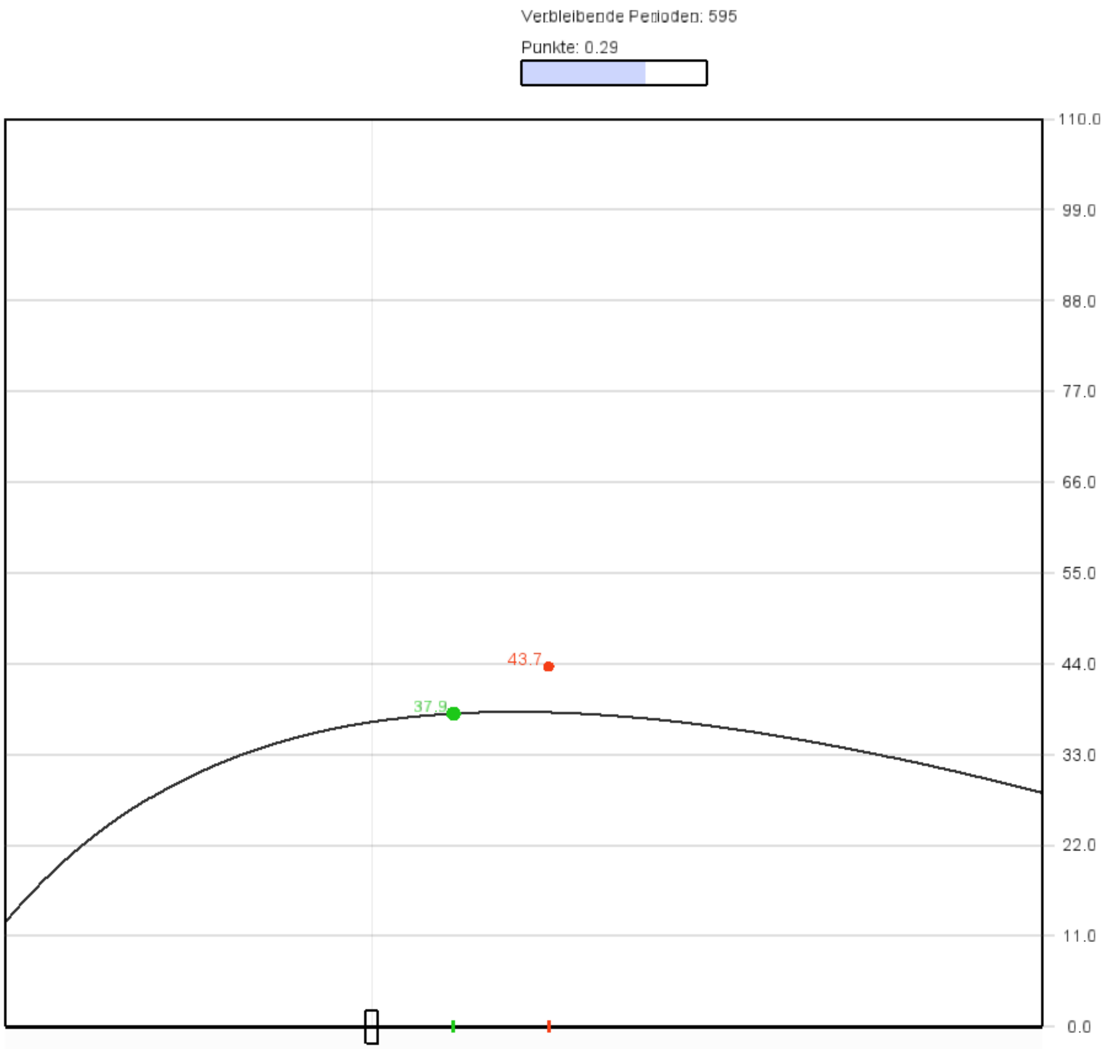


Figure 1: Interface in the Info treatment

The subject chooses her action on the horizontal axis via a slider. The green and red dots represent a subject's, respectively her partner's, earnings in the preceding period. The black line shows the potential payoffs in the preceding period. On the top of the interface subjects see the number of remaining periods, the amount of points they earned, and how much time they have left to decide via a progress bar. The interface in NoInfo is exactly the same except for the black curve of potential payoffs. Note that, in the experiment, the x-axis is intentionally left unlabelled to avoid giving the subjects any cues on how to behave. The y-axis shows the payoffs in points.

volatile and stressful than the sheer number of periods might suggest.

The experiment was run employing the ConG software package (Pettit et al., 2014). In each of the two treatments we observe 18 independent anonymous pairs of subjects who play the game for 600 periods. Subjects were recruited via ORSEE (Greiner, 2015) at the WZB-TU laboratory in Berlin. Sessions lasted 80 minutes and subjects' average payouts were 12.63 Euro in Info and 15.66 Euro in NoInfo, plus a 5 Euro show-up fee. Upon arrival subjects were randomly allocated to seats and received written instructions explaining (i) they would play against one other participant; (ii) how they would choose from their action spaces; and (iii) what feedback they would receive through the graphical interface. Subjects are not made aware of the payoff function. They only know that the payoff function depends exclusively on both players' choices and that it is symmetric and does not change over time. The experimental instructions are reported in appendix A.

1.2.2 Aggregate Results

Figure 2 shows median actions over the course of the experiment. The experimental results can be divided into three phases: First a phase in which subjects employ naïve heuristics shaped by salient feedback information; second a learning phase in which subjects revise their initial naïve heuristics; and, third, a long-run phase where behaviour settles down. In what follows we separate the phases at period 25 and period 300. All results are robust to changing the separating periods by $\pm 20\%$.

Information provided has a strong influence on which naïve heuristics subjects initially choose. In the naïve NoInfo heuristics phase subjects choose actions above the Nash prediction - an often-observed consequences of imitate-the-best heuristics that subjects frequently employ in low-information settings in previous experiments. (Friedman et al., 2015; Huck et al., 1999; Offerman et al., 2002; Apesteguia et al., 2007;

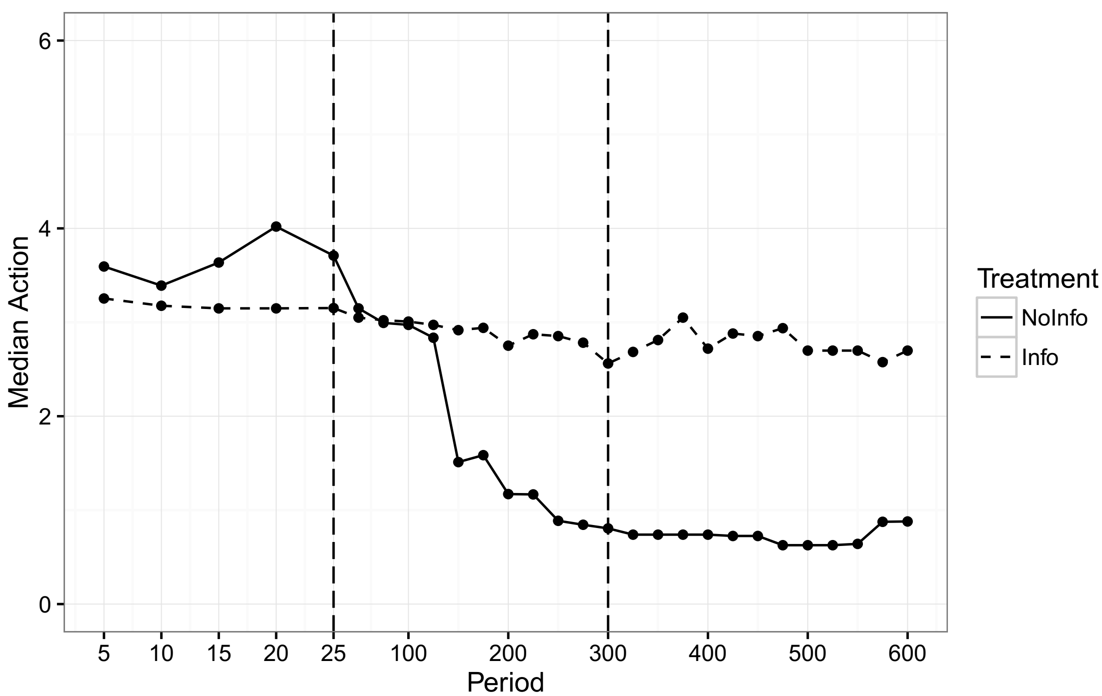


Figure 2: Median Actions

Up to period 25 each dot represents the median action in bins of 5 periods. From period 25 (dashed vertical line) on each dot represents the median action in bins of 25 periods. In the naïve phase (periods 1-25) the average action (standard deviation in parentheses) is 3.65 (1.58) in NoInfo and 3.37 (1.19) in Info. In the learning phase (periods 26-300) this changes to 2.22 (1.90) in NoInfo and 2.93 (1.65) in Info. Finally, in the long-run phase (periods 301-600) behaviour settles down at 1.23 (1.45) in NoInfo and 2.62 (1.80) in Info.

Boylan and El-Gamal, 1993). In contrast, during the same initial naïve phase, subjects in the Info treatment rapidly converge on the static Nash equilibrium outcome. This behaviour is consistent with a myopic best-reply heuristic, also documented in previous studies in which payoff information was available (Friedman et al., 2015; Offerman et al., 2002; Apesteguia et al., 2007; Cheung and Friedman, 1997). The difference between the two treatments is significant (two-sided MWU test, $p < 0.05$).

After 25 periods, subjects in NoInfo dramatically decrease their actions, indicating that they successfully abandon the imitation heuristic. Subjects in Info exhibit a much weaker downward pattern, suggesting that they have a harder time abandoning the myopic best-reply heuristic. From period 300 on behaviour stabilizes in the long-run phase. In NoInfo subjects choose actions that are significantly more cooperative than the Nash outcome (two-sided t-test, $p < 0.001$) whereas in Info their average action is indistinguishable from the Nash outcome (two-sided t-test, $p > 0.1$). Crucially, subjects cooperate much more successfully in treatment NoInfo than in Info (two-sided MWU test, $p < 0.001$), generating also significantly higher payoffs in the long-run phase. In the final phase median earnings are 6.01 Euro in Info whereas in NoInfo they are over 50% higher at 9.56 Euro. Note that this pattern is replicated in the treatments with 1,200 and 2,400 periods, though subjects in the Info treatments show slightly more collusive behaviour. See figures 17 and 18 as well as the discussion in appendix A.

1.2.3 Individual Behaviour

What individual behavioural patterns give rise to the differences in aggregate behaviour documented above? To find out, we run simple regressions to investigate if the changes that we observe are consistent with a set of heuristics relevant in our setting. Specifically, we examine "Imitate-the-Best" which tells a subject to copy the opponent's action in the previous period if she chose a higher action and made higher profits; "Match" which tells a subject to copy the other player's action regardless of payoff; "Win-Continue, Lose-Reverse" (WCLR) where a subject repeats the previous round's adjustment if it improved her payoff and changes directions if profits declined; and, finally, "Myopic Best Reply" which tells subjects to move towards the best reply against her opponent's current action. While imitate-the-best pushes behaviour towards extremely competitive outcomes (a higher action always generates a higher relative payoff) (Vega-Redondo, 1997) and myopic best replies push behaviour towards Nash (Hofbauer and Sandholm, 2002), both, the Match heuristic and WCLR are heuristics that foster cooperation. As with tit-for-tat, Match rewards an opponent's move towards cooperative play and punishes deviations to higher actions. Moreover, Match aligns subjects' actions thereby speeding up WCLR as an effective force for pushing actions towards cooperation (Axelrod and Hamilton, 1981; Huck et al., 2003, 2004). Obviously, Myopic Best Reply is easily available for subjects in the Info treatment whereas it requires information acquisition and much more sophistication in the NoInfo treatment.

Some of these heuristics give point predictions whereas others yield predictions about the direction of change. To make comparable comparisons, we focus only on changes in decisions relative to the subject's previous period action. The outcome variable CHANGE gives the direction of change. Independent variables COPY UP and COPY DOWN capture the Imitate-the-Best and Match heuristics. Finally WCLR is captured by the variable of the same name and Myopic Best Reply is captured by variable BR. The treatment effects are captured by treatment dummy INFO. We also interact treatment dummy INFO with all of the heuristics variables described above in order to pick up the differences in subjects' use of the modelled heuristics. We define the variables in the regression in table 1 as follows. The out-

	Dependent Variable CHANGE	
	(1) Periods 1-25	(2) Periods 26-600
COPY UP	0.422***	0.274***
	(0.144)	(0.077)
COPY DOWN	0.015	-0.221**
	(0.172)	(0.082)
WCLR	0.062*	0.108**
	(0.031)	(0.049)
BR	0.114***	0.087*
	(0.037)	(0.047)
INFO	0.355*	-0.196*
	(0.186)	(0.109)
COPY UP:INFO	-0.518***	0.124
	(0.168)	(0.106)
COPY DOWN:INFO	-0.362*	0.285**
	(0.196)	(0.109)
WCLR:INFO	-0.006	-0.065
	(0.053)	(0.057)
BR:INFO	0.226***	0.111*
	(0.066)	(0.062)
Constant	0.264	0.347***
	(0.161)	(0.082)
Observations	1,428	26,216
R-squared	0.233	0.250
F-Statistic	47.91***	81.48***

Standard errors clustered on groups in parentheses.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 1: Linear Probability Model

come variable CHANGE is coded as 1 for an increase and 0 for a decrease in action. Imitate-the-best is captured by the COPY UP dummy which is coded as 1 if the opponent chose a higher action in the previous period. The match heuristic is covered by, both, COPY UP and COPY DOWN (which is coded as 1 if the opponent had a lower action). WCLR is captured by the dummy of the same name which is coded as 1 if the heuristic prescribes an increase and 0 otherwise. For constant profits, we code the variable as 1. The results are robust to coding it as 0 or dropping these observations. Finally, myopic best-replies are covered by the BR dummy which is coded as 1 when the myopic best reply was above a subject's previous period action and 0 otherwise. There are no observations for which the myopic best-reply would be exactly zero. Finally, we have a treatment dummy INFO which is 1 for the Info treatment and 0 for NoInfo. We report a linear probability model estimated by OLS as the interpretation of the coefficients, especially of the interaction terms, is most straightforward. The results of a Logit estimation are qualitatively similar. To check robustness we also run regressions where we drop observations on the limit of the action space as subjects can then only increase respectively decrease their action. The results remain qualitatively robust. Variable definitions are reported in the Methods section. Note that interaction terms are denoted by a colon. In this regression we pool data from learning and long-run phases as results do not change qualitatively between them. Table 13 in appendix A reports separate regressions for learning and long-run phases.

Table 1 reports the results of the model estimated separately on the naïve, and then the learning and long-run phases combined. The first column reports estimates for the naïve phase, i.e. the first 25 periods. The coefficient COPY UP is significant and large whereas COPY DOWN is insignificant. This indicates strong imitate-the-best behaviour in NoInfo. In Info on the other hand there is no evidence of imitate-the-best behaviour. The interaction term COPY UP:INFO is negative and weakly significant and cancels out the baseline effect entirely (Wald test $p > 0.1$). In NoInfo the BR coefficient picks up some regression to the mean while the large and highly significant coefficient for the interaction term BR:INFO indicates just how appealing subjects find the best-reply heuristic once they have the relevant information. Finally, there is some weakly significant evidence for WCLR already in the first phase. Some subjects may already switch from imitation to WCLR at the end of the initial phase.

Column 2 reports estimates for the learning and the long-run phases (periods 25 to 600). In NoInfo COPY UP remains significant while COPY DOWN now also turns significant and negative. There is therefore a switch from Imitate-the-best to the Matching heuristic. The estimations also show that WCLR becomes a key heuristic in NoInfo. This behaviour gives rise to the decrease in actions in the learning phase and stabilizes cooperative play in the long-run. The interaction term WCLR:INFO is negative if insignificant, and the linear combination of baseline and interaction term is insignificant (Wald test $p > 0.1$). Thus there is no evidence of subjects following WCLR in the Info treatment. Instead the interaction effect BR:INFO is weakly significant and when combined with the baseline is highly significant (Wald test $p < 0.01$), indicating that subjects continue strongly to follow the Myopic Best Reply heuristic. These results thus show that subjects find it much harder to abandon their initially chosen heuristic in Info.

1.3 Discussion

Payoff information can hurt because it can affect the initial heuristics subjects become attached to as well as subjects' long-run affinity to that heuristic. This is in stark contrast to rational models of cooperation. The results strongly suggest that subjects lean heavily on heuristics instead of employing more sophisticated repeated-game logic to achieve cooperation in early play as much as in the long run. Most importantly, dropping superficially plausible heuristics like myopic best reply (which maximizes short-run payoffs) turns out to be much harder than dropping a more obviously ill-adapted heuristic like imitate-the-best (which generates much lower payoffs by inducing extreme competition). As a result, payoff information is an advantage only in the short run, hampering the learning necessary to establish cooperation in the long run. Modelling such long-run evolution of cooperation requires a two-tier model where subjects follow heuristics for a while but adapt or replace them over time. Fully accounting for such long-run cooperative evolutionary processes requires models that describe not only the heuristics employed by agents in an interaction but also how and when agents learn to improve these heuristics over time. Finally, some thoughts on external validity are in order. In our laboratory setting feedback on others' actions and profits and feedback on possible payoffs that could have been achieved are extremely salient and there are no information channels that would connect individuals to others outside their immediate group of competitors. We would, of course, expect that learning reacts to changes in the environment such as readily available multi-period memory, information from other groups of players, or endogenous choice of partners and evolutionary selection. Indeed, incorporating such richer information and social structures provides many interesting possibilities for future research. However, we believe that when it comes to the external validity of our main result we can argue that our Info treatment provides an ideal setting for payoff information to aid subjects' rationality. There is very little distraction and experimentation to learn more about the payoff landscape remains, thanks to the large number of repetitions, extremely cheap. If human cooperation were largely driven by rational calculation our Info treatment should clearly have speeded up cooperation and should have increased subjects' earnings throughout the experiment. We find evidence to the contrary despite the ideal conditions provided in the laboratory. Our findings make it unlikely that individual payoff information, that is, readily available feedback on one's own performance and advice on how one could have done even better would be of any help to foster cooperation in richer environments closer to real life.

2 Learning and Reciprocity in Adjustments

Abstract

I propose a notion of reciprocity in adjustments of actions rather than – as commonly assumed in the literature – actions. I reanalyze experimental data from chapter 1 to investigate the role of reciprocity in adjustments for learning and cooperation. I find that reciprocity plays a significant role in learning but reciprocity does not eliminate win-continue lose-reverse (WCLR) as an important cooperation-supporting heuristic. Additionally, I find strong gender effects. Female and mixed gender groups are less likely to find cooperative agreements. Analyzing individual behavior, I find that this is driven by women who are less likely to reciprocate kind actions.

2.1 Introduction

In the previous chapter we find evidence for a win-continue lose-reverse (WCLR) (Huck et al., 2003) learning process in the NoInfo treatment after subjects abandon their initial imitation heuristic. WCLR is a process in which agents who choose actions from an ordered set first go into one direction, either increasing or decreasing their choice. If their profits increase they continue in the same direction. If their profits decrease, though, they turn around. Central to the WCLR process in a Cournot environment is that, as both agents decrease their actions both, their payoffs increase thus encouraging them to stay on course to lower, more cooperative actions.

There is another possible explanation for such downward adjustment in actions. Note that in such a strategic environment lowering one's action is essentially a kind action, even more so below the symmetric stage-game best reply of 3. Below the best reply, lowering one's action decreases one's own payoff while increasing the partner's payoff by more than one's own decrease. When following WCLR, both agents move towards lower actions simultaneously and thus both gain. Two kind actions in one period thus encourage more kind actions in the next period. Such a behavioral pattern is reminiscent of reciprocity. Note that the previous literature on reciprocity differs from this notion of reciprocity in one central aspect. Typically, economic models of reciprocity evaluate how kind an action is. I propose reciprocity in adjustments of actions instead and define an adjustment that increases the partner's payoff as kind. Reciprocity has been studied much in economics. For a summary of the early literature on reciprocity in economics see Rabin (1993) and Fehr and Gächter (2000). Reciprocity has been used to explain gift exchange in competitive environments in the laboratory (Fehr et al., 1998) as well as in the field (Falk, 2007). Reciprocity is accepted as a basic behavioral force in human behavior (Bowles and Gintis, 2002; Fehr and Gächter, 1998; Dohmen et al., 2009). Furthermore, theoretical models have been proposed to understand the effects of reciprocity in strategic environments (Cox et al., 2007; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006).

The workhorse game for measuring reciprocity is the trust game (Berg et al., 1995). Typically, a sender is endowed with some money that she can send to her partner. The amount that is sent is tripled by the experimenter. The receiver then can share some of her funds with the sender. Sending back money is interpreted as reciprocating the kind action of trust displayed by the receiver. In a review of the literature Croson and Gneezy (2009) find that, as receivers, women tend to reciprocate more than men but the picture is not very clear. Some studies find no gender differences, and Bellemare and Kröger (2007) even find men to be more reciprocal than women. Garbarino and Slonim (2009) argue that gender differences in reciprocating behavior depend on age and the amount received in a complex manner. In a large-scale experiment representative of the German population, Dittrich (2015) finds men to be more reciprocating than women. These results suggest that there may be gender differences in reciprocating behavior, but it is not clear in which direction the effects may go.

Learning to cooperate has also been theoretically studied in theoretical biology. Frean (1996), Roberts and Sherratt (1998), Wahl and Nowak (1999b) and Wahl and Nowak (1999a) look at the continuous prisoners' dilemma, a game with similarities to the Cournot game we study. Importantly, Wahl and Nowak (1999b) identify strategies that let agents learn to cooperate. Such strategies are characterized by generosity. Agents following such a strategy follow up their partner's action with a slightly more cooperative action. Two agents following such a strategy move on a trajectory towards full cooperation.

The question I am going to address in this chapter is whether the competing hypotheses of WCLR and reciprocity can be distinguished. Note that while WCLR and reciprocity can both explain a downward trajectory towards more cooperative choices there are subtle differences. First, they differ in mechanism. For reciprocity, all that matters is the partner's action. If one's partner was kind and decreased her action, reciprocity predicts a decrease in one's own action as well. If on the other hand the partner was not kind and increased her action, reciprocity would predict an increase in one's own action afterwards. For WCLR the predictions are more nuanced. WCLR would predict a decrease in one's action if one previously decreased one's action and one's profits increased at the same time. But if one decreased one's action by too much and one's profits subsequently decreased, WCLR would predict an increase in action even if one's partner was nice and decreased her action.

Interestingly, I find that both WCLR and reciprocity play a role in the subjects' path to cooperation in the NoInfo treatments. When WCLR and reciprocity pull in the same direction, most subjects follow the path prescribed by those heuristics. When they prescribe different behavior, actions can go in either direction. After testing these competing hypotheses, I go deeper into the data and explore the heterogeneity in actions. There is a lot of variance in outcomes in our data. Even in treatment NoInfo, in which a lot of groups manage to find a cooperative outcome, there are some groups who do not manage to cooperate. I find strong gender effects on cooperation. Female groups and mixed gender groups are a lot less likely to cooperate than male groups. The evidence suggests that women are less likely to reciprocate kind actions, and thus are less likely to find a cooperative agreement.

2.2 Reciprocity in Adjustments

Differently than the previous literature on reciprocity, I consider kindness and reciprocity in adjustments of actions rather than in actions themselves. In our experiment subjects have a continuous action set from which they can choose their action in period t $x_{i,t} \in [0.1, 6]$. The payoff in period t depends on a subject's own action $x_{i,t}$ and her partner's action $x_{-i,t}$. Specifically, we implement the payoff function $\pi_{i,t}(x_{i,t}, x_{-i,t}) = 10 + (\frac{120}{\sum_j x_{j,t}} - 10)x_{i,t}$.

$$\frac{d\pi_{i,t}}{dx_{-i,t}} = -\frac{120x_{i,t}}{(\sum_j x_{j,t})^2}$$

Note that $\pi_{i,t}$ decreases in $x_{-i,t}$ for $x_i > 0$. A decrease in one's own action strictly increases the partner's payoff. Hence, I define a decrease in action $\Delta x_{i,t} = x_{i,t} - x_{i,t-1} < 0$ to be kind. If reciprocity plays a role in subjects' choices I expect subjects to repay a kind action with a kind action and an unkind action with an unkind one. After observing a step towards more cooperative behavior by the partner $\Delta x_{-i,t} < 0$ a subject should answer with a step towards more cooperative behavior herself in the following period $\Delta x_{i,t+1} < 0$. Similarly, if the partner behaved unkind $\Delta x_{-i,t} > 0$ a subject would reciprocate and move towards less cooperation in the

Observed Change in			Predicted Adjustment	
Partner's Action	Own Action	Own Payoff	Reciprocity	WCLR
$\Delta x_{-i,t} < 0$	$\Delta x_{i,t} < 0$	$\Delta \pi_{i,t} < 0$	$\Delta x_{i,t+1} < 0$	$\Delta x_{i,t+1} > 0$
	$\Delta x_{i,t} < 0$	$\Delta \pi_{i,t} > 0$	$\Delta x_{i,t+1} < 0$	$\Delta x_{i,t+1} < 0$
	$\Delta x_{i,t} > 0$	$\Delta \pi_{i,t} < 0$	$\Delta x_{i,t+1} < 0$	$\Delta x_{i,t+1} < 0$
	$\Delta x_{i,t} > 0$	$\Delta \pi_{i,t} > 0$	$\Delta x_{i,t+1} < 0$	$\Delta x_{i,t+1} > 0$
$\Delta x_{-i,t} > 0$	$\Delta x_{i,t} < 0$	$\Delta \pi_{i,t} < 0$	$\Delta x_{i,t+1} > 0$	$\Delta x_{i,t+1} > 0$
	$\Delta x_{i,t} < 0$	$\Delta \pi_{i,t} > 0$	$\Delta x_{i,t+1} > 0$	$\Delta x_{i,t+1} < 0$
	$\Delta x_{i,t} > 0$	$\Delta \pi_{i,t} < 0$	$\Delta x_{i,t+1} > 0$	$\Delta x_{i,t+1} < 0$
	$\Delta x_{i,t} > 0$	$\Delta \pi_{i,t} > 0$	$\Delta x_{i,t+1} > 0$	$\Delta x_{i,t+1} > 0$

Table 2: Comparing Predictions of WCLR and Reciprocity

following period $\Delta x_{i,t+1} > 0$. If the partner didn't change her action I assume the subject to increase, decrease or not change her action with equal probability. Formally, given some step size β and a random variable drawn from an i.i.d uniform discrete distribution $z_t \sim U\{-1, 0, 1\}$

$$x_{i,t+1} = x_{i,t} + \beta \begin{cases} sign(\Delta x_{-i,t}) & \text{if} \quad \Delta x_{-i,t} \neq 0 \\ z_t & else \end{cases} \tag{1}$$

How does this rule compare to WCLR? For WCLR the subject should only care about her own change in actions and about her partner's change in actions only indirectly if her partner's change affects her own payoff $\Delta \pi_{i,t} = \pi_{i,t} - \pi_{i,t-1}$. Formally,

$$x_{i,t+1} = x_{i,t} + \beta \begin{cases} sign(\Delta x_{i,t}) * sign(\Delta \pi_{i,t}) & \text{if} \quad sign(\Delta x_{i,t}) * sign(\Delta \pi_{i,t}) \neq 0 \\ z_t & else \end{cases} \tag{2}$$

Table 2 compares the predictions of WCLR and reciprocity. Note that some lines are more likely than others. If the partner decreased her action in the previous period it is very likely that one's own payoff increased.

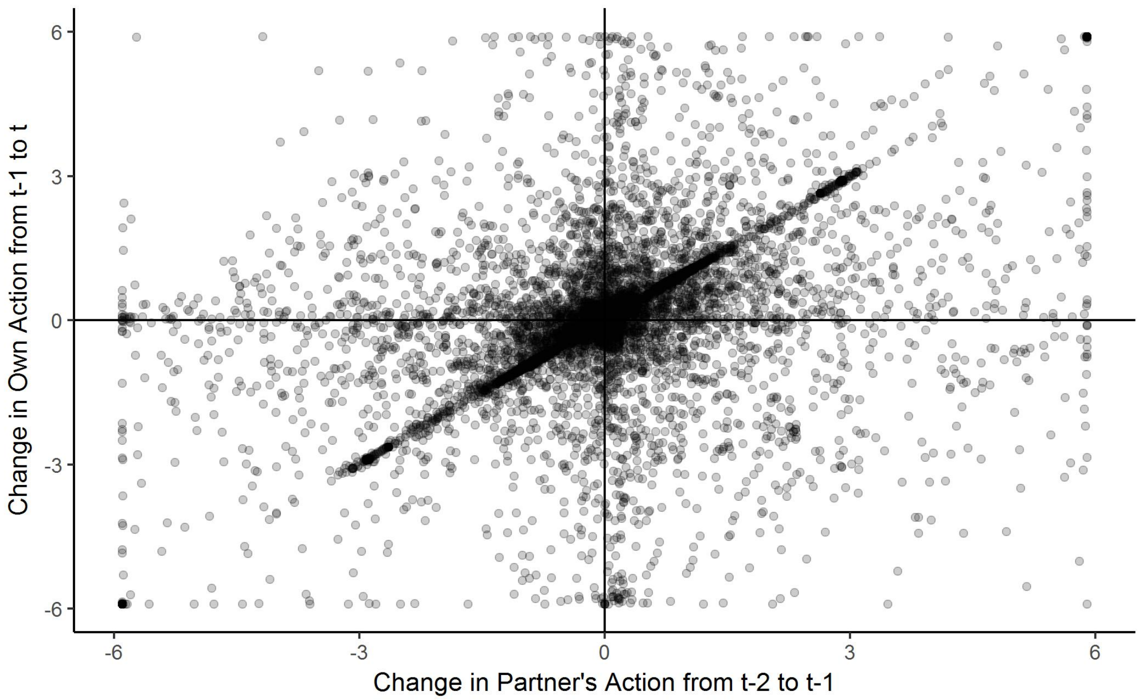


Figure 3: Scatterplot of Change in Partner's Action in Previous Period in Relation to Change in Own Action for 8 Seconds NoInfo Treatment

Can we find identify a pattern of reciprocity in the data? Figure 3 shows how subjects react to a change in their partner's action in the 8 seconds NoInfo treatment. This is the treatment in which we found most cooperative behavior in the previous chapter. The plot clearly shows a positive relationship between the partner's previous change in action and the subsequent change in the subject's own action, indicating that reciprocity might play a role here. The bigger the increase in action of her partner, the bigger the increase in action of the subject in the next period.

But figure 3 cannot tell the entire story as information necessary for evaluating WCLR as a subject's change in action and her change in profits are missing. Figure 4 splits figure 3 up into four subplots. In the first two panels in the top row, the predictions for WCLR and reciprocity coincide. In the first subplot, the red dots show subjects' actions after having decreased their action and having earned lower

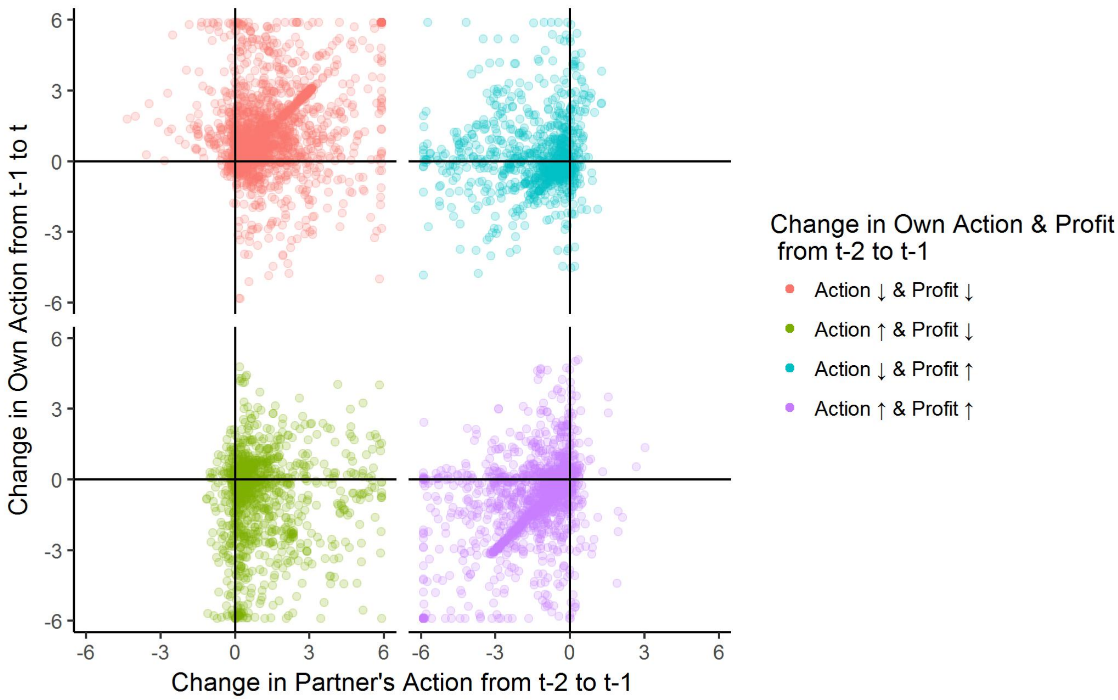


Figure 4: Scatterplots of Change in Partner's Action in Previous Period in Relation to Change in Own Action for 8 Seconds NoInfo Treatment, by Own Change in Action in Previous Period and Change in Own Profits

profits than before. This is mostly due to the subject's partner having increased her action. WCLR predicts that the subject should turn around and increase her action. Reciprocity predicts the subject to reciprocate an unkind with an unkind action and thus increase her action as well. Thus, both WCLR and reciprocity predict an increase in action, which the data clearly shows. The second panel with blue dots shows subjects' behavior after having decreased their action and having increased profits. This is mostly due to the subject's partner having decreased her action as well. Thus, both WCLR and reciprocity predict a decrease in action, which the data seems to support.

Finally, in the two panels in the bottom row the predictions for reciprocity and WCLR differ. In the third panel with green dots, I plot subjects' behavior after having increased their action and subsequently earning less than before. This is mostly due to the partner having increased her action as well, obviously an unkind action. Thus, reciprocity would predict a further increase in action. On the other hand, WCLR predicts a decrease in action, as the increase had bad consequences for payoffs. The data shows actions going in both directions, ruling out neither explanation. A similar pattern arises in the fourth panel, which shows subjects behavior after having increased their action and having earned more. This is mostly due to their partner having decreased her action at the same time. In this case, WCLR would predict a further increase in action while reciprocity would predict a decrease. In the data, there are many observations in either direction.

What can we learn from these plots? It seems as if reciprocity and WCLR both play a role here. If both pull in the same direction, as the panels in the top row of figure 4 show, subjects mostly follow their predictions. But if they pull in different directions, subjects seem to follow neither fully. A regression similar to the one in table 1 in chapter 1 confirms this intuition, as table 3 shows. In line with the previous chapter I only consider subjects' actions after the initial learning period of 200 seconds, i.e. 25 periods in the 8 seconds treatment, 50 periods in the 4 seconds treatment, and 100 periods in the 2 seconds treatment. I choose a linear probability model to make the interpretation of the coefficients easier. Results for a logit regression are qualitatively similar and are reported in appendix B in table 15. Furthermore, due to length of the tables some coefficients have been cut. The full regression can be found in appendix B in table 16. The dependent variable is a dummy that is 1 if a subject increased her action, and 0 else. COPY UP is a dummy that is 1 if the partner chose a higher action than oneself in the previous period, and 0 else. COPY DOWN is a dummy that is 1 if the partner chose a lower action than oneself in the previous period, and 0 else. WCLR is a dummy that is 1 if the WCLR heuristic would predict an increase in action, and 0 else. BR is a dummy that is 1 if the best response in the previous period, given the partner's action, would have been higher, and 0 else. INFO is a dummy that is 1 if the subject is in the Info treatment group, and 0 else. NICE OWN (t-1) is a dummy that is 1 if a subject was kind in the previous period, i.e. if she chose a lower action in period t-1 than in t-2, and 0 else. NICE OTH (t-1) is a dummy that is 1 if the partner was kind in the previous period. Similarly, NICE OWN (t-2) and NICE OTH (t-2) show if the subject or her partner were nice two periods ago. NICE BOTH (t-1) is a dummy that is 1 if both the subject and her partner were nice in the last period. INCPROFIT is a dummy that is 1 if the subject's profits increased in the last period. Finally, a colon indicates an interaction term.

Columns M1 through M3 show results for the 8 seconds treatment only to make comparisons with chapter 1 easier. The first column M1 replicates the regression from our previous study. We interpreted these results as subjects following a heuristic that matches their partner's action plus a WCLR heuristic leading them towards cooperation. Column M2 adds how kind subjects acted from period t-2 to period t-1. NICE OTH (t-1) is negative and highly statistically significant whereas NICE OWN (t-1) is insignificant. Subjects seem to reciprocate their partner's kind behavior in the last period by also decreasing their own action. One's own nice behavior in the last period does not have any significant impact. An interesting pattern arises when looking at kind behavior from two periods before in column M3. Now NICE OWN (t-2) is negative and highly significant, whereas NICE OTH (t-2) is insignificant. Thus, if a subject behaved kind two periods ago she is very likely to be kind in the current period but her partner's behavior two periods ago does not seem to matter. Taking these effects together a story of reciprocity emerges. Subjects reciprocate each other's kind behavior from period to period and thus get on a path towards

	Dependent Variable: CHANGE				
	8 Seconds Treatment			All Treatments	
VARIABLES	M1	M2	M3	M4	M5
COPY UP	0.274***	0.243***	0.328***	0.253***	0.254***
	(0.0769)	(0.0707)	(0.0656)	(0.0367)	(0.0368)
COPY DOWN	-0.221**	-0.144*	-0.0557	-0.00611	-0.00657
	(0.0821)	(0.0811)	(0.0650)	(0.0399)	(0.0402)
WCLR	0.108**	0.128***	0.137***	0.0801***	
	(0.0485)	(0.0407)	(0.0371)	(0.0133)	
BR	0.0871*	0.0878**	0.0696**	0.0920***	0.120***
	(0.0466)	(0.0393)	(0.0314)	(0.0164)	(0.0171)
INFO	-0.196*	-0.191*	-0.0970	-0.0881*	-0.0908*
	(0.109)	(0.101)	(0.0908)	(0.0504)	(0.0508)
COPY UP:INFO	0.124	0.140	0.0370	0.0724	0.0731
	(0.106)	(0.0999)	(0.0924)	(0.0553)	(0.0561)
COPY DOWN:INFO	0.285**	0.266**	0.165	0.0689	0.0712
	(0.109)	(0.106)	(0.0975)	(0.0582)	(0.0593)
WCLR:INFO	-0.0651	-0.0663	-0.0645	-0.00541	
	(0.0573)	(0.0499)	(0.0448)	(0.0179)	
BR:INFO	0.111*	0.102*	0.103**	0.0192	0.0161
	(0.0617)	(0.0556)	(0.0495)	(0.0290)	(0.0278)
NICE OWN (t-1)		0.0662	0.0523	-0.0503**	-0.0287
		(0.0392)	(0.0378)	(0.0213)	(0.0208)
NICE OTH (t-1)		-0.195***	-0.144***	-0.162***	-0.123***
		(0.0345)	(0.0287)	(0.0158)	(0.0178)
NICE OWN (t-2)			-0.179***	-0.223***	-0.221***
			(0.0218)	(0.0155)	(0.0155)
NICE OTH (t-2)			-0.0413*	0.000105	0.00224
			(0.0227)	(0.0150)	(0.0153)
NICE BOTH (t-1)					0.0128
					(0.0273)
NICE BOTH (t-1):					-0.113***
INCPROFIT					(0.0219)
Constant	0.347***	0.362***	0.368***	0.432***	0.448***
	(0.0818)	(0.0773)	(0.0574)	(0.0386)	(0.0382)
Time Treatment Controls	n.a.	n.a.	n.a.	Y	Y
Observations	26,216	26,216	23,005	113,698	113,698
R-squared	0.249	0.285	0.317	0.234	0.231

Standard errors clustered on groups in parentheses.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 3: OLS Regression on Direction of Change

cooperation. These effects are robust to including data from other treatments with 2 and 4 seconds per period as column M4 shows. The full model M4 shows that if a subject's partner was kind she is about 14 % more likely to be kind as well and decrease her action. If a subject was kind two periods ago she is 18 % more likely to be kind again. In other words, subjects get locked on a mutually beneficial reciprocal path towards cooperation. WCLR is also a significant factor. If the heuristic would prescribe an increase or decrease the subject is 8 % more likely to follow.

Hence, the positive results for reciprocal learning behavior do not eliminate WCLR. Going from M1 to M4 and adding dummies for kind behavior does not change the results for WCLR. The results remain statistically significant and relevant in size. Remember that WCLR is quite a complicated variable though. Its value depends directly on the change in one's action and, through the change in one's own payoff, indirectly on the change the partner's action from period t-2 to period t-1. Since the focus lies on subjects learning to cooperate I report a robustness check in column M5. In this regression, I introduce two more variables while dropping WCLR. NICE BOTH (t-1) is an interaction term between a subject's own and her partner's kind action dummy in the last period. NICE BOTH (t-1):INCPROFIT interacts this term with INCPROFIT which is 1 if one's profits increased in the previous period. This term catches the WCLR heuristic on its downward trajectory. If reciprocity were the entire story both variables should be insignificant. Instead, I find NICE BOTH (t-1):INCPROFIT to be negative and highly significant. So, if both subjects in the same market were nice in the period before a subject decreases her action only if her profits increased as well in the last period. This effect is comparable in size with the WCLR effects in models M1 through M4.

As a robustness check I simulate two data sets and run regressions similar to model M4. In one data set agents follow WCLR, in the other one they follow a reciprocity process. When agents follow WCLR the WCLR variable is significant but the reciprocity variable is insignificant. When agents follow reciprocity, the picture is reversed. Now the WCLR variable is insignificant while the reciprocity variable is significant. The simulation procedures are documented in appendix B and the the regression results are reported in table 19 in the same appendix. The results show that my chosen regression model can distinguish between WCLR and reciprocity in the data.

Thus, both learning through reciprocity as well as learning through WCLR seem to be present in the data. It is conceivable that subjects follow both heuristics, and sometimes follow one or the other. Alternatively, it could be that some subjects follow WCLR whereas some follow a reciprocity trajectory. More experiments would be required to tease these explanations apart.

2.3 Heterogeneity: Cognitive Skills and Gender

A closer look at our data reveals much heterogeneity between groups. There is a wide range of outcomes within treatments. Even though cooperation is much more prevalent in the NoInfo treatments there are groups that do not manage to cooperate. Figure 5 pools the results from all time treatments (2, 4, and 8 seconds) and plots the average group choice over the course of the entire experiment. The treatment effect of giving subjects more payoff information is striking, as the average action in Info is around the one-shot Nash equilibrium action of 3 while in NoInfo it is much lower. Still, there is considerable variance in outcomes.

What could be the reason for such variance between groups? On the one hand, if subjects in the experiment differ in their propensity to cooperate then by chance some groups would end up cooperating more than others. Previous research suggests

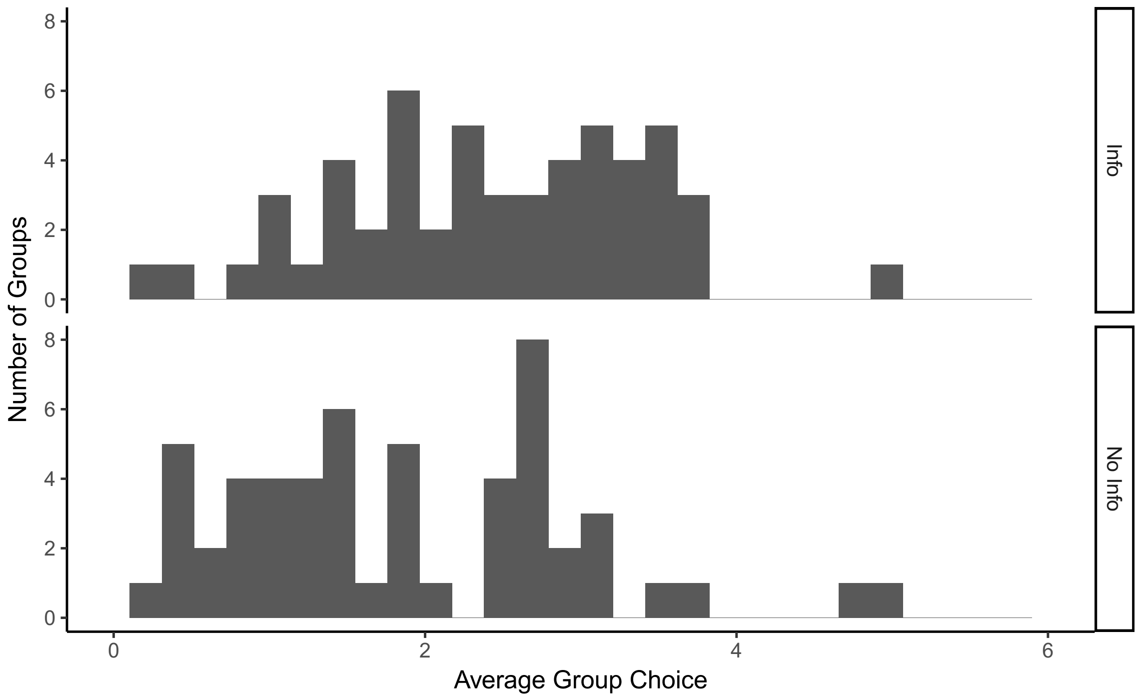


Figure 5: Heterogeneity in Average Group Choices, All Time Treatments Pooled

that cooperativeness may correlate with gender (see for example Croson and Gneezy (2009)) so the gender composition of groups should be considered. Do purely male groups behave differently from purely female or mixed groups? On the other hand, choosing, trying out and abandoning badly performing heuristics is an intellectual task so another possibility may be that subjects' intellectual characteristics play a role. High-ability subjects could be expected to abandon ill-chosen heuristics more easily than low-ability subjects. If two high-ability subjects are matched to each other then we could expect them to learn more easily and arrive at a cooperative outcome more quickly than low-ability subjects. In the questionnaire, we ask subjects their gender, their self-rated math ability on a scale of 1 to 10 and what their mother tongue is. I divide the groups using these characteristics. First, by gender into male groups, female groups, and mixed groups. Second, by math ability into groups with both subjects being above average in their math ability, both being below, and mixed groups. Third, by mother tongue into groups with both subjects having German as their mother tongue, and groups with at least one subject not having German as their mother tongue.[1]

Over all sessions and treatments, 32 groups are male, 26 female, and 50 are of mixed gender. 17 groups are below average in self-reported math ability, 43 are above average, and 17 are mixed.[2] Finally, 68 groups have two German native speakers and 40 have at least one non-German native speaker.

Figure 6 plots the average action in bins of 100 second and splits the graphs by group characteristics. The first panel splits the sample by the number of females, the second panel by the number of non-native German speakers, and the third panel by the number of subjects stating a math-ability above average in a group. I pool all the time treatments to get enough observations per group category. The number of non-native German speakers does not seem to play a role as the average actions are very similar across groups. For self-rated math ability and gender there are differences between groups. It seems as if two subjects with self-rated math ability above average perform better than subjects with self-rated math ability below average in both Info and NoInfo. For gender, in the Info treatment there seems to be no difference in action over time. Male groups, female groups, and mixed groups behave very similarly. In the NoInfo treatment though, there is a sizeable gender gap. Initially, all groups start out with highly competitive actions but male groups quite quickly move towards more cooperative actions. Female groups on average have a higher action at every single point in time, with mixed gender groups being in between. Earlier I documented that

[1]Note that out of 108 groups there is only one group with no German native speaker.

[2]The average self-rated math ability is 6.63. Note that there are more subjects above average in ability since self-rated math ability is an integer value between 1 and 10 but the average is a real number. As a thought experiment, consider a group of 4 subjects. When asked about their math ability, three subjects state 10 and one subject states 2. The average math ability would be 8. Three subjects would be above that average, one below.

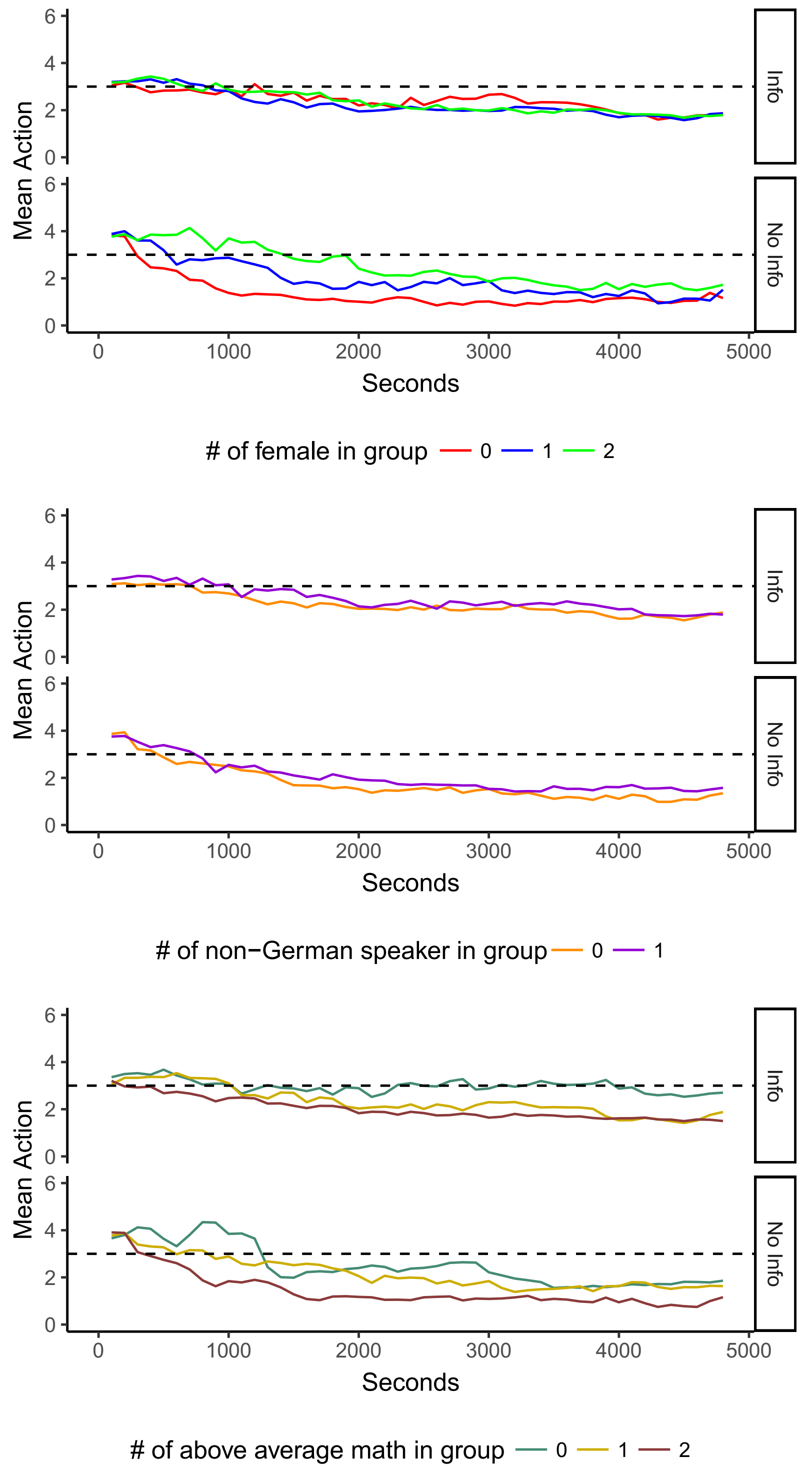


Figure 6: Average Action in Bins of 100 Seconds, by Group Characteristics, Pooling All Treatments

the learning process in NoInfo is driven by reciprocity and WCLR. Since we observe a gender gap in NoInfo but not in Info I would expect a gender gap in reciprocity or WCLR. I will use regressions to identify which characteristics explain these differences in behavior best.

Of course, this analysis is post hoc. Initially, we did not design the experiment with these questions in mind. Still, group composition has been exogenously and randomly imposed. If one were to run an experiment on group composition one would employ similar procedures as we did in the initial experiment. The only change might be an invitation procedure to guarantee balanced groups. Therefore, I can look at group characteristics and study their causal impact on behavior. Still, I am testing multiple hypotheses. Multiple hypothesis testing implies that, by chance, some results will be statistically significant although the relationship is spurious. See List et al. (2016) for a recent take on the issue. Since I am testing multiple hypotheses without an ex-ante hypothesis I correct for multiple testing by adjusting and tightening the significance levels. In the regressions that follow I test five post hoc variables. Hence, in the statistical analysis that follows I adjust the significance levels using a conservative Bonferroni correction of $\frac{\alpha}{5}$.

But before going into individual behavior let us go deeper into the data. At the individual level, the heterogeneity becomes even more striking. Consider the following illustrative example. Figure 7 shows two groups picked from the same treatment NoInfo 8 seconds. Group 1 displays behavior which is in line with our explanation in the previous study. Initially, subjects choose very high and competitive actions. Over time they learn to lower their actions until they end up at the joint profit maximum. There are two short deviations from the joint profit maximum as one subject seems to test if her opponent is still paying attention but after her opponent reacts quickly they return to cooperation. Only at the very end of the game does this arrangement

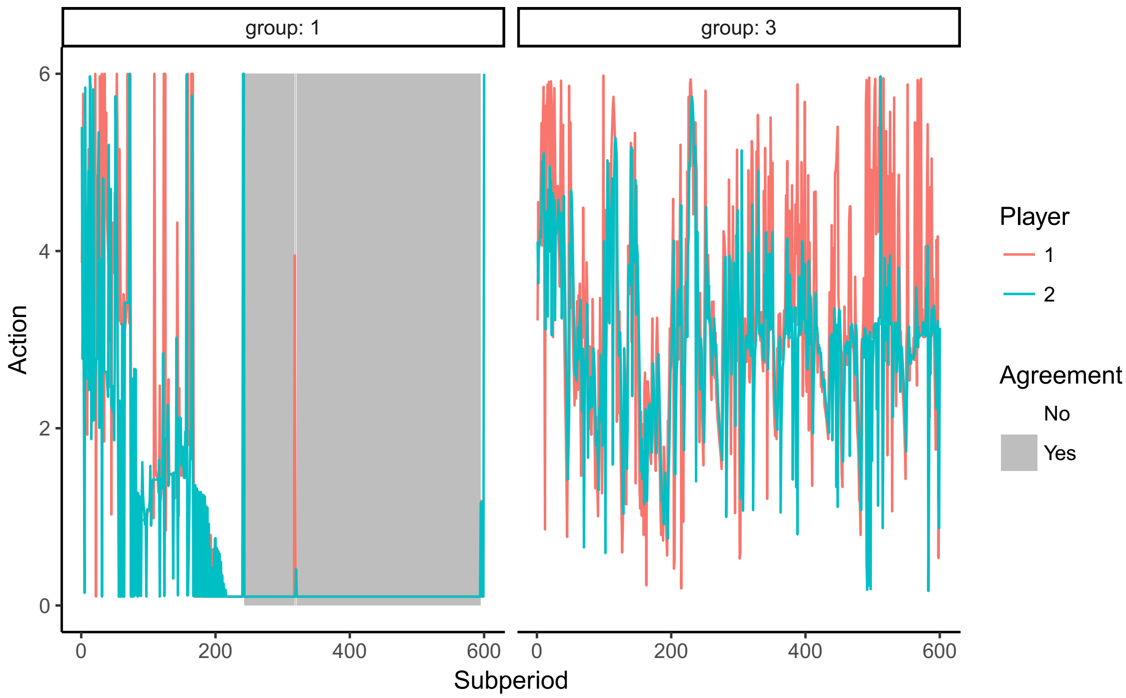


Figure 7: Two Groups in 8 Seconds NoInfo

unravel. Compare this group's behavior to group 3's behavior. Initially, these subjects also start out with highly competitive actions but do not manage to form a stable agreement for the entirety of the experiment.

I distinguish between groups who are in a stable agreement and groups who are not. When two subjects do both not change their actions for at least 300 seconds (or 5 minutes) I define them to be in a stable agreement. Note that all the following results are robust to both decreasing or increasing the time window by up to an order of magnitude and to considering actions within a small band as a stable arrangement. 52 % of groups achieve at least one stable agreement. Figure 7 illustrates my classification. Roughly from period 200 onwards, group 1 enters a stable agreement which is two times broken for a short time. But after a short punishment phase subjects agree again on cooperation. Group 3, on the other hand, never finds an agreement.

Using this distinction, I can investigate two questions. First, conditional on having found an agreement, is there a significant impact of group characteristics on the average action that the subjects agree on? Second, is there a significant influence of group characteristics on the probability of finding an agreement? Table 4 reports a regression of group characteristics on the average action in an agreement, conditional on the group having found one. Note that subjects in an agreement are a very selected sample. The results therefore should not be interpreted as the effects of group characteristics on the average action but as the effects on the average action *conditional* on having found an agreement. Significance levels have been adjusted for multiple testing. M/F GROUP is a dummy that is 1 if the group is a mixed gender group, and 0 else. F/F is a dummy that is 1 if the group is a female group, and 0 else. MIXED MATH is a dummy that is 1 if the group has one below average and one above average math ability subject. GOOD MATH is a dummy that is 1 if both subjects are above average in math ability. MIXED LANG is a dummy that is 1 if there is at least one non-German native speaker in the group. 4 SEC and 8 SEC are dummies for groups in the 4 seconds and 8 seconds treatments. INFO is a dummy for groups in the Info treatment.

Groups with subjects who self-report a high math ability agree on lower actions, but not significantly so. For mother tongue and gender there is no significant impact on the average action in an agreement either. Hence, I find no significant relationship between group characteristics and the average action subjects in a group can agree on.

But what about the probability of finding an agreement? Table 5 reports the results of a regression on finding an agreement.[3] The dependent variable is a dummy that is 1 if the group finds at least one agreement during the 80 minutes of the

[3]Again, I report a linear probability model to make the interpretation of the coefficients easier. In appendix B in table 14, I report a logit model that yields qualitatively similar results.

VARIABLES	AVERAGE ACTION IN AGREEMENT
M/F GROUP	-0.29
	(0.40)
F/F GROUP	-0.12
	(0.59)
MIXED MATH	-0.49
	(0.66)
GOOD MATH	-1.05
	(0.64)
MIXED LANG	0.35
	(0.38)
4 SEC	-0.31
	(0.42)
8 SEC	-0.69
	(0.49)
INFO	0.42
	(0.38)
Constant	1.86**
	(0.68)
R-squared	0.15
Observations	56

Standard errors in parentheses
$^{*}p < 0.02, ^{**}p < 0.01, ^{***}p < 0.002$

Table 4: OLS Regression on Average Action in Group, Conditional on Being in Agreement

VARIABLES	AGREEMENT
M/F GROUP	-0.18*
	(0.10)
F/F GROUP	-0.33**
	(0.12)
MIXED MATH	0.02
	(0.13)
GOOD MATH	0.15
	(0.14)
MIXED LANG	-0.07
	(0.09)
4 SEC	-0.27**
	(0.11)
8 SEC	-0.43***
	(0.11)
INFO	-0.14
	(0.09)
Constant	0.94***
	(0.16)
R-squared	0.27
Observations	108

Standard errors in parentheses
$^{*}p < 0.02, ^{**}p < 0.01, ^{***}p < 0.002$

Table 5: OLS Regression on Finding an Agreement

experiment. The independent variables are the same as in table 4. Again, significance levels have been adjusted for multiple testing. This time, there are strong gender effects. Female groups are a lot less likely to find an agreement than male groups. With female groups being about 33 % less likely to agree the effect is economically large and significant even when accounting for multiple testing. Mixed gender groups are about 18 % less likely to find an agreement than male groups, but only marginally significantly so. Math ability and language does not have a significant impact on the probability of the group finding an agreement.

Interestingly, the effect for female groups is about twice as large as the effect for mixed gender groups. Therefore, I look closer at individual behavior to investigate how women might behave differently from men in the experiment.

I adapt model M4 from table 3, interacting the heuristics with a FEMALE dummy that is 1 if the subject is female. Table 6 reports the results.[4] I separate the treatments into two regressions to avoid too many interaction terms. Column M6 reports the regression for the NoInfo treatments. Strikingly, there is only one significant interaction term. Women are about 9 % less likely to reciprocate a kind action from their opponent. This is quite a large decrease from the 23,5 % chance of men reciprocating a kind action from their opponent. Thus, the gender gap in actions that I observe in the NoInfo treatments can be explained by women reciprocating kind actions less. Interestingly this gender difference disappears in column M7, which reports the same regression for the Info treatments. Now there is no significant gender difference in reciprocity, which aligns nicely with there being no gender gap in actions in the Info treatments. It seems as if Info levels the playing field in this regard.

[4]Some coefficients have been cut due to length of the table. The full table can be found in appendix B in table 17. I report a linear probability model to make the interpretation of the coefficients easier. In appendix B in table 18, I report a logit model that yields qualitatively similar results.

	Dependent Variable CHANGE	
	NoInfo	Info
VARIABLES	M6	M7
COPY UP	0.242***	0.388***
	(0.0711)	(0.0444)
COPY DOWN	-0.0473	0.109**
	(0.0778)	(0.0490)
WCLR	0.0767***	0.0536***
	(0.0148)	(0.0137)
BR	0.0660***	0.130***
	(0.0209)	(0.0248)
NICE OWN (t-1)	-0.0559**	-0.0832**
	(0.0237)	(0.0364)
NICE OTH (t-1)	-0.235***	-0.141***
	(0.0270)	(0.0266)
NICE OWN (t-2)	-0.217***	-0.223***
	(0.0198)	(0.0313)
NICE OTH (t-2)	0.0295	-0.00504
	(0.0200)	(0.0292)
FEMALE	-0.0643	0.0943
	(0.0869)	(0.0667)
COPY UP:FEMALE	0.00491	-0.0955
	(0.0795)	(0.0678)
COPY DOWN:FEMALE	0.0659	-0.0788
	(0.0858)	(0.0691)
WCLR:FEMALE	0.0160	0.0344
	(0.0154)	(0.0222)
BR:FEMALE	0.0333	-0.0268
	(0.0278)	(0.0371)
NICE OWN (t-1):FEMALE	0.00804	0.0583*
	(0.0445)	(0.0314)
NICE OTH (t-1):FEMALE	0.0886**	-0.0110
	(0.0363)	(0.0293)
NICE OWN (t-2):FEMALE	-0.0200	0.0149
	(0.0278)	(0.0278)
NICE OTH (t-2):FEMALE	-0.0114	-0.0303
	(0.0242)	(0.0215)
Constant	0.487***	0.277***
	(0.0717)	(0.0509)
Time Treatment Controls	Y	Y
Observations	51,970	61,728
R-squared	0.251	0.226

Standard errors clustered on groups in parentheses.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 6: OLS Regression on Direction of Change with Gender Dummies

2.4 Discussion

This paper contributes to different subfields in experimental and behavioral economics. First, I introduce a new idea of reciprocity to the literature. Rather than considering absolute levels of actions, reciprocity in adjustments looks at the *changes* in actions from one period to the next. In our experiment a decrease in actions is unambiguously kind and I find that subjects reciprocate such a kind action with a decrease in their own action in the following period. One might think of WCLR and reciprocity as two competing hypotheses in our Cournot framework. Both explain subjects' movement towards more cooperative actions until subjects reach a cooperative agreement. Since their mechanisms are different, I can use our existing data to tease these explanations apart. The evidence suggests that subjects are both following WCLR and reciprocity. Often, WCLR and reciprocity pull subjects in the same direction towards lower actions, and subjects then tend to follow the learning rules' predicted actions closely. In some cases though, WCLR and reciprocity pull in different directions. In these cases I observe subjects going in both directions. These results suggest that both reciprocity and WCLR play a significant role in explaining subjects' behavior.

Second, this paper is related to the literature on heterogeneity of social preferences. The literature suggests that there may be gender differences in reciprocity, with some studies suggesting women to be more reciprocal (see Croson and Gneezy (2009) and references therein) and others finding women to be less reciprocal than men (Bellemare and Kröger, 2007; Dittrich, 2015). Going deeper into individual behavior, I find significant gender differences. Female and mixed gender groups are significantly less likely to find a cooperative agreement, even after accounting for multiple testing. Using the insights from the first section on reciprocity and learning, I find that women are significantly less likely to reciprocate a kind action. It could be that subject pools are heterogeneous in what kind of men and women are represented in them. Different universities specialize in different fields, which may attract some kind of women or men more than others. This study contributes a further piece of evidence to the growing literature on behavioral gender differences.

Future work may extend the work on reciprocity in adjustments and formulate a full model. Many economic relationships evolve over time without the partners in such a relationship fully cooperating from the get go. Reciprocity in adjustments may allow the partners to slowly learn to trust each other more and let them move towards more and more cooperative outcomes.

Angabe zum bereits veröffentlichtem Kapitel

Das folgende Kapitel ist die Preprint-Version eines Artikels, der bereits als Discussion Paper in der Reihe der WZB Discussion Papers veröffentlicht wurde. Am 16. März 2018 wurde der Artikel bei der Zeitschrift Games and Economic Behavior (Elsevier Verlag) eingereicht.

Ensthaler, Ludwig, Steffen Huck und Johannes Leutgeb (2017): Games played through agents in the laboratory – A test of Prat & Rustichini's model. WZB Discussion Papers, Research Unit: Economics of Change SP II 2016-305r, Social Science Research Center Berlin (WZB). https://bibliothek.wzb.eu/pdf/2017/ii16-305r.pdf

Statement on previously published chapter

The following chapter is the preprint version of an article which was previously published as a discussion paper in the WZB Discussion Papers series. On 16 March 2018 the article was submitted to the journal Games and Economic Behavior (publisher Elsevier).

Ensthaler, Ludwig, Steffen Huck and Johannes Leutgeb (2017): Games played through agents in the laboratory – A test of Prat & Rustichini's model. WZB Discussion Papers, Research Unit: Economics of Change SP II 2016-305r, Social Science Research Center Berlin (WZB). https://bibliothek.wzb.eu/pdf/2017/ii16-305r.pdf

3 Games Played Through Agents in the Laboratory

Abstract

From the regulation of sports to lawmaking in parliament, in many situations one group of people ("agents") make decisions that affect payoffs of others ("principals") who may offer action-contingent transfers in order to sway the agents' decisions. Prat and Rustichini (2003) characterize pure-strategy equilibria of such *Games Played Through Agents*. Specifically, they predict the equilibrium outcome in pure strategies to be efficient. We test the theory in a series of experimental treatments with human principals and computerized agents. The theory predicts remarkably well which actions and outcomes are implemented but subjects' transfer offers deviate systematically from equilibrium. We show how quantal response equilibrium accounts for the deviations and test its predictions out of sample. Our results show that quantal response equilibrium is particularly well suited for explaining behavior in such games.

3.1 Introduction

Consider an individual, A, whose payoff depends on how some agents play a game. This individual has no immediate choice to make in this game, yet its outcome might be of considerable importance to her. Imagine a pharma company whose business opportunities depend on a new piece of legislation to be voted on in parliament. Or a motor racing team whose competitive edge might depend on new rules to be introduced by the governing body of the sport. Or, perhaps, a child who has no direct say in her parents' relocation decision or choice of holiday destination but passionately cares about both. In all these cases A will mull over the consequences of the game played by the agents and her only option to influence the outcome is through some form of "bribery", that is, through providing incentives to the agents, who play the game, to take particular choices that may lead to one of her more desired outcomes.

Of course, at the same time, there might be another individual, B, who also cares about the outcome of this game and her incentives might differ from A's. B will then compete with A for influence over the agents. The situation that arises is essentially a bundle of auctions where A and B vie for influence by promising payments to the agents for making particular choices. A might offer a payment to some member of parliament for voting against the new legislation, while B might offer a payment to the same parliamentarian for voting for the new law. Of course, the parliamentarian might also have intrinsic preferences over the different outcomes that can ensue. Her total utility will then be a function of both, her intrinsic utility and the payments she can ensure for herself by voting for one of the actions that A and B have incentivized.

The full game that arises from this structure has the following timing. First, A and B make binding payment promises to all the agents. Then the agents play the game. This interesting and important class of games has been introduced by Prat and Rustichini (2003) and termed *Games Played Through Agents* or GPTAs. We will stick to this terminology and will refer below to the players who play the actual game as the *agents* while we will refer to A and B as the *principals*. GPTAs are solved through backward induction. Under the assumption of additively separable utility, Prat and Rustichini (2003) study the equilibrium solutions for this class of games and find, somewhat surprisingly, that subgame perfect equilibria are characterized by strong efficiency. Specifically, equilibria yield efficient outcomes that maximize the sum of all players' payoffs, principals and agents alike.

Despite their wide applicability GPTAs have, to the best of our knowledge, not yet been subjected to an empirical investigation. Our study provides a first attempt to explore the empirical validity of Prat and Rustichini's (2003) theory by means of a laboratory experiment. As the equilibrium logic is far from being trivial we focus on some of the most basic cases, all based on 2×2 games where the agents who play the game have no intrinsic preferences over the four possible outcomes (perhaps such as some politicians who mainly value the income they obtain from lobbyists). Moreover, in order to avoid complications stemming from social preferences that could muddle the relation between agents and principals, we computerize the agents who simply implement the action that maximizes their income.

We study four types of GPTAs, where the payoff matrices, from the viewpoint of the principals, resemble the following games: a prisoners' dilemma (PD), a coordination game (COORD), a simple dominance solvable game (DOM) and finally a battle of the sexes (BoS). The PD and COORD games have symmetric payoff matrices while the DOM and BoS ones are asymmetric. For the PD we implement two different payoff structures to test one of the theory's key comparative statics which, incidentally, also includes predictions for the coordination game. While we find the equilibrium logic for the symmetric games already quite demanding, the asymmetric games are even harder to solve due to the inherent conflict between principals.

Analyzing the data from our experiment we find that the theory does remarkably well. In each of our games the predicted actions are implemented most of the time and the comparative statics for equilibrium offers hold empirically across games and are highly significant. We did not necessarily expect this to be the case and were surprised by the theory's predictive power. Of course, the question arises how good is the fit really? What is the comparison? The point prediction of all probability mass concentrated on one outcome is, of course, falsified. So how can we say that the theory is doing a really good job organizing the data? On the one hand, we appeal to standard notions of good fit in experimental economics where hitting the target 80 %

of the time tends to be a success. On the other hand, we examine the systematic deviations in offers from equilibrium predictions in more detail. We find a strong asymmetry in deviations: principals' offers are close to equilibrium for actions that matter most to them while offers for actions that are typically not implemented are too low. This is intuitive and can be shown to influence our subjects' reasoning from the very beginning of the game. In order to quantify these deviations in a systemic manner, we estimate a structural logit quantal response equilibrium (QRE) model. Our precision estimates are within the usual range and, for the symmetric games, the fit between QRE predictions and data is almost scarily good. To test the *predictive* power of our QRE framework out of sample we implement a further treatment, DOM. In the symmetric treatments QRE predicts an asymmetry *within* a principal's offers but in DOM, QRE predicts an asymmetry *between* two principals' offers. Our data exhibits the predicted pattern, thus supporting the QRE approach. For the asymmetric BoS the QRE's fit is still very good but there are systematic divergences between the model predictions and the data that have its origin in the more challenging underlying equilibrium logic. Our results suggest that subjects violate the agent indifference condition in Prat and Rustichini's (2003) theory. Integrating QRE into the original theory allows us to explain the deviations and also successfully predict behavior out of sample.

In all, our paper shows that despite its non-trivial logic, Prat and Rustichini's (2003) theory does succeed in the laboratory. This should encourage, both, further empirical studies of GPTAs (in the laboratory but perhaps also in the field) as well as more applications of the theory which so far have been rare.[5]

[5] The applications that we are aware of have dealt with lobbying in a federal (Bordignon et al., 2008) or international setting (Aidt and Hwang, 2008), where lobbyists take the role of principals and states or nations the role of agents. Fredriksson and Millimet (2007) apply Prat and Rustichini's (2003) theory to a game of pollution taxation and lobbying and argue that macro-level patterns in gasoline prices are consistent with the theory.

The remainder of the paper is organized as follows. Section 3.2 summarizes the key insights of Prat and Rustichini (2003). In section 3.3 we introduce the symmetric games, discuss the experimental results and show how quantal response equilibrium organizes the data. In section 3.4 we move on to the asymmetric games to test QRE and the theory further. Finally, section 3.5 discusses the results and concludes.

3.2 Games Played Through Agents: A Very Short Summary

A *Game Played Through Agents* or GPTA, as defined in Prat and Rustichini (2003), is played by a set of agents who have to choose actions and a set of principals who offer each agent a schedule of monetary transfers contingent on chosen actions. An agent chooses her action to maximize the sum of transfers she receives from the principals plus any intrinsic cost or benefit of her choice. A principal chooses her transfer schedule to maximize her utility from the agents' actions minus the sum of the transfers she makes to the agents. A GPTA is modeled as a two-stage game where, first, the principals simultaneously choose their transfer schedules and, second, all agents simultaneously choose their actions. Prat and Rustichini (2003) derive necessary and sufficient conditions for the existence of a pure-strategy equilibrium in these games with arbitrary numbers of agents, principals, and actions.

In this paper we only study games with two agents and two principals. Furthermore, each agent only has two actions to choose from and the agents derive no intrinsic utility from an outcome, that is, their payoff from an action simply equals the sum of the promised transfers for taking that action. Let the game's notation be as follows. An agent $n \in N = \{R, C\}$ chooses an action $s_n \in S_n$, where $S_R = \{U, D\}$ and $S_C = \{L, R\}$. We will refer to the agents as row agent and column agent. The combination of agents' actions translates into an outcome

	L	R
U	π^A_{UL}, π^B_{UL}	π^A_{UR}, π^B_{UR}
D	π^A_{DL}, π^B_{DL}	π^A_{DR}, π^B_{DR}

Table 7: Generic Principals' Payoff Matrix

$s \in S = \prod_{n \in N} S_n = \{UL, UR, DL, DR\}$. The principals $m \in M = \{A, B\}$ receive payoffs depending on the outcome. We will refer to the principals as principal A and principal B. Let π^m_s denote the gross payoff to principal m for outcome s. The principals' payoffs can then be represented in a matrix like table 7. Principals engage in a bidding game and offer transfers contingent on actions to agents. Let $t^m_n(s_n)$ denote the transfer principal m offers to agent n for choosing action s_n. If the agents implement outcome s principal m receives the net payoff $\pi^m_s - \sum_{n \in N} t^m_n(s_n)$. Agents do not derive intrinsic utility from the different outcomes of the game and, thus, simply choose the action that implies the highest total transfer. Framing the problem differently, the principals engage in two simultaneous sealed-bid first-price auctions where their valuation for winning each auction depends on whether they win or lose the other auction.[6]

Prat and Rustichini (2003) derive a particularly simple equilibrium characterization for this class of games. They show that the following three conditions characterize a pure-strategy equilibrium $(\hat{t}, \hat{s})$: Agent Indifference (AI), Incentive Compatibility (IC), and Cost Minimization (CM).

$$\sum_{m \in M} \hat{t}^m_n(\hat{s}_n) = \sum_{m \in M} \hat{t}^m_n(s'_n) \ \forall s'_n \in S_n, \forall n \in N \ \textbf{(AI)}$$

[6] For the case of two principals and two agents who implement equilibria in a 2×2 subgame, that is, the case we study in the laboratory, GPTAs coincide with Bikhchandani (1999)'s auctions of heterogeneous objects. For larger GPTAs this equivalence does not hold as in a GPTA principals' valuations do not only depend on the set of auctions that they win but also on the identity of the winner of those auctions that they do not win themselves.

Agent Indifference says that agents are indifferent between the equilibrium action $\hat{s}_n$ and the alternative s'_n. Transfer offers are the same for both actions as otherwise some principal could decrease her transfer offer and implement the same outcome. Notice that while agents are indifferent, in equilibrium, they will implement the efficient outcome as otherwise the equilibrium logic would unravel.

$$\pi^m_{s'} - \pi^m_{\hat{s}} \leq \sum_{n \in N} \hat{t}^{j \neq m}_n(\hat{s}_n) - \sum_{n \in N} \hat{t}^{j \neq m}_n(s'_n) \ \forall m \in M, \forall s' \in S \ \textbf{(IC)}$$

Incentive Compatibility states that the gains from implementing a different outcome are outweighed by the costs of incentivizing the agents to do so. On the left-hand side of the inequality is the gain in gross payoffs for principal m from moving from $\hat{s}$ to s'. On the right-hand side is the difference in the other principal's total offers for the actions that implement $\hat{s}$ and s' respectively, or in other words, the additional costs that have to be borne in order to incentivize the agents to switch to another outcome. Remember that the matrix shown in table 7 only shows the principals' payoffs that result from the 2×2 game *played by the agents*—it is *not* the game played by the principals. In a GPTA, a principal must also compare payoffs moving along the diagonal (or off-diagonal) as she has to compare losing both auctions to winning both of them.

$$\hat{t}^m_n(\hat{s}_n) > 0 \Rightarrow \hat{t}^m_n(s'_n) = 0 \ \forall m \in M, \forall n \in N \ \textbf{(CM)}$$

Finally, Cost Minimization requires that if a principal offers a positive transfer for the equilibrium action $\hat{s}_n$, she will not do so for the alternative s'_n. Else a principal could decrease her offers for both actions and still implement the same outcome. From these conditions it is easy to see that the equilibrium outcome must be efficient in the strong sense that in equilibrium the *sum* of payoffs to principals and agents will be maximal. Simply, sum (IC) over all principals and by (AI) the right-hand side equals zero. Therefore, the sum of gains from implementing a different outcome than $\hat{s}$ cannot be positive, hence, the equilibrium outcome maximizes the sum of payoffs. In the following, we will refer to this simply as the efficient outcome.[7]

$$\sum_{m\in M} \pi^m_{s'} - \sum_{m\in M} \pi^m_{\hat{s}} \leq \sum_{m\in M}\sum_{n\in N} \hat{t}^m_n(\hat{s}_n) - \sum_{m\in M}\sum_{n\in N} \hat{t}^m_n(s'_n) = 0$$

3.3 The Experiment: Symmetric Settings

We implement three symmetric GPTAs. The first game, PD High, uses a prisoners' dilemma payoff matrix with a high temptation payoff. The second game, PD Low, does the same but with a lower temptation payoff. Finally, the third game, COORD, uses a coordination game payoff matrix, in which the principals' incentives are perfectly aligned. Table 8 summarizes the payoff matrices for principals in the three treatments. In order to avoid negative payoffs we also introduce budget constraints equal to the maximum gross payoff that is attainable. Therefore, in PD High each principal's budget is 6, in PD Low it is 5, and in COORD it is 3. In Prat and Rustichini's (2003) model principals can offer transfers for all the agents' actions but will, in equilibrium, only incentivize one of each. For the purposes of the experiment, we simplify the setup by restricting principals to offer transfers only for one of each agent's actions, namely the one they prefer. In the symmetric treatments we restrict principal A to offer transfers for L and D, and conversely principal B may offer transfers for R and U. So $t^A_R(D), t^A_C(L), t^B_R(U), t^B_C(R) \geq 0$ and $t^A_R(U), t^A_C(R), t^B_R(D), t^B_R(L) = 0$.

[7] The analysis in Prat and Rustichini (2003) extends to more general classes of games. Also, they derive results for the case where the transfers of the principals are not contingent on the actions of the individual agents but on the outcome of the game which is a result of the actions of all agents. We do not discuss these results here as they are beyond the scope of our experimental analysis.

	L	R
U	$4,4$	$0,6$
D	$6,0$	$1,1$

PD High

	L	R
U	$4,4$	$0,5$
D	$5,0$	$1,1$

PD Low

	L	R
U	$3,3$	$0,0$
D	$0,0$	$1,1$

COORD

Table 8: Symmetric Principals' Payoff Matrices

Consequently, **(CM)** is satisfied by design in all treatments.

PD High If a pure-strategy equilibrium exists it must have UL as an outcome, as this is the unique efficient outcome. By **(AI)**, the equilibrium offers are of the form $t_R^A(D) = t_R^B(U)$ and $t_C^A(L) = t_C^B(R)$ and hence both principals make both agents indifferent between choosing either of their actions. Restating the incentive constraints **IC** we know that

$$4 + t_C^A(L) \geq \ max\{t_R^A(D) + t_C^A(L), 6, 1 + t_R^A(D)\},$$

and

$$4 + t_R^B(U) \leq max\{t_R^B(U) + t_C^B(R), 6, 1 + t_C^B(R)\}$$

which simplifies to

$$2 \leq t_R^A(D) = t_R^B(U) \leq 4 \text{ and } 2 \leq t_C^A(L) = t_C^B(R) \leq 4.$$

As both agents have a budget of 6, we arrive at the following set of pure-strategy equilibria with UL as the outcome for PD High.

$$(t_R^A(D) = t_R^B(U), t_C^A(L) = t_C^B(R)) \in \{(a, b) \mid a, b \in [2, 4] \wedge a + b \leq 6\}$$

Intuitively, think of principal A who considers an outcome UL that promises her a gross payoff of 4. First, if the row agent were to choose D and thus implement DL principal A would receive a gross payoff of 6 for a net gain of 2. Thus, as long as principal B offers at least 2, principal A has no incentive to offer more. Instead she can match principal B's offer as, in equilibrium, the agent will choose U and she will not have to pay her offer. Second, consider that principal A were to decrease her offer to the column agent and would consequently lose the auction for this agent. Her gross payoff would then fall by 4. Therefore, as long as principal B offers at most 4 to the column agent, principal A will offer the exact same amount and will have to pay the offer for the column agent. Third, in order to implement DR she would have to lose the auction for L while winning the auction for D. From the first two cases we know that from losing the former she would gain at most 4 while winning the latter would cost at least 2. The net gains in offers would amount to at most 2, while the net loss in gross payoffs would be 3. Thus, the conditions from the first two cases imply the conditions for the third. Fourth, by symmetry, the same considerations apply to principal B. Therefore equilibrium offers must be between 2 and 4. We shall refer to

$$t_R^A(U) = t_R^B(D) = t_C^A(L) = t_C^B(R) = 2$$

as the (unique) lowest-transfer equilibrium.

PD Low By the same reasoning as in PD High we find that the set of pure-strategy equilibria with UL as the outcome for PD Low is

$$(t_R^A(U) = t_R^B(D), t_C^A(L) = t_C^B(R)) \in \{(a, b) \mid a, b \in [1, 4] \wedge a + b \leq 5\}$$

Again, there exists a unique lowest-transfer equilibrium where

$$t_R^A(U) = t_R^B(D) = t_C^A(L) = t_C^B(R) = 1$$

COORD As in PD High and PD Low, UL is the only efficient outcome and hence must be the outcome in any pure-strategy equilibrium. The set of pure-strategy equilibria with UL as the outcome is

$$(t_R^A(U) = t_R^B(D), t_C^A(L) = t_C^B(R)) \in \{(a, b) \mid a, b \leq 0 \wedge a + b \leq 2\}$$

The unique lowest-transfer equilibrium in this game is

$$t_R^A(U) = t_R^B(D) = t_C^A(L) = t_C^B(R) = 0$$

3.3.1 Experimental Implementation and Hypotheses

In order to reduce complexity and focus on the most interesting part of the model we computerize the two agents in all treatments. This also eliminates potential effects from social preferences between agents and principals which is realistic in many real-life settings where principals and agents might come from very different walks of life. As all games that we implement are 2×2 games, one agent is always called the row agent and the other agent the column agent. Moreover, for the symmetric games discussed above we chose to have identical instructions for all participants. That is, every principal received instructions as if they were principal A matched up against a principal B. When matching two principals, the computer then transformed one of the principal's offers into that of an equivalent principal B. Hence, offers for U and L (D and R) are equivalent and we can pool them into U/L (D/R).

Subjects are assigned the role of one of the two principals. Each subject receives an initial budget equal to the maximum payoff in the payoff matrix. The subjects can use this budget to incentivize the agents. In order to simplify matters further principals can only incentivize one of the two actions for each agent. (Remember that, in equilibrium, it would never make sense to offer an agent payments for both possible actions.) For example, principal A can only incentivize the row agent to choose D and the column agent to choose L. Agents will then choose the action for which they are offered the higher amount. In case both principals offer an agent the same amount, we need, of course, a tie-braking rule in the experiment. In the theoretical model such a tie-braking rule is not necessary. As discussed above, the theory predicts the outcome that maximizes the sum of payoffs. Yet, agents will be indifferent between their two actions in equilibrium. However, the equilibrium logic forces them to implement the efficient outcome as, if they did not, the equilibrium would unravel. For the experiment it appeared unwise to say that the computerized agents would follow "equilibrium logic." Hence, we decided instead to have an explicit tie-breaker. The tie-breaking rule simply states that if the column agent is offered the same amount of money by both principals she will choose L and that if the row agent is offered the same amount of money by both agents, she will choose U.

Let us now derive some hypotheses for play in the three symmetric games that we expect to hold in the experiment. First, we would expect subjects to post offers such that the agents implement U and L.

Hypothesis 1. *U and L are the modal implemented actions in PD High, PD Low and COORD.*

Next notice that the offers for $(U/L, D/R)$ in the unique lowest-transfer equilibrium decrease from PD High to PD Low to COORD, from $(2, 2)$ to $(1, 1)$ to $(0, 0)$, respectively. Hence, from a comparative static perspective we would expect to see offers decrease from PD High to PD Low to COORD.

Hypothesis 2. *Median offers decrease from PD High to PD Low to COORD.*

Finally, all equilibria predict symmetric offers.

Hypothesis 3. *Principals choose symmetric offers in each of the three games.*

Procedures, player matching and all other details of the implementation are the same in treatments PD High, PD Low, and COORD and the instructions are fully analogous.[8] There are 30 rounds of the same game in every session and matching between rounds is random. After each round subjects receive feedback about the offers of their opponent, the resulting outcome and their net earnings. The first two rounds are practice rounds while the remaining 28 rounds are payoff relevant. At the end of the experiment, two rounds out of those are chosen with equal probability and the earnings in those rounds are paid out to subjects. For each symmetric treatment we run two experimental sessions with 24 subjects each. Subjects are randomly assigned to matching groups of eight subjects. Hence, there are 48 subjects in six disjoint matching groups for every treatment. Subjects are matched within these matching groups for the duration of the experiment. In every round, a subject is randomly assigned to one of the other seven subjects in her matching group, throughout the experiment with the same uniform probability. Before the start of the first round subjects have to take an understanding test (see appendix C which also contain the instructions). Subjects are only allowed to continue with the experiment after answering all questions in the test correctly. Altogether, we run five treatments, with the three symmetric games described above as well as two asymmetric setups described further below. The sessions were run between 2014 and 2016 at the WZB-TU

[8]The translated instructions for PD High are provided in appendix C as an example.

laboratory in Berlin with subjects who are undergraduate or master's students at one of the major research universities in town. In total 264 subjects participated in the experiment. The experiment was implemented in z-Tree (Fischbacher, 2007) and subjects were recruited for the experiment using ORSEE (Greiner, 2015). The points that subjects earned were paid out 1:1 in Euro. In all treatments subjects received a participation fee of 5 Euro which was paid out regardless of their choices in the experiment. On top of the participation fee subjects earned 13.5 Euro on average. At the end of the experiment subjects were paid out in cash in private. 87 % of subjects were German nationals, 58 % of subjects were male and average age was 24.3. Subjects had been studying for 5.1 semesters on average and they reported to study a wide range of subjects with 8 % studying economics, with 75 % of subjects having had at least one math module during their studies.

3.3.2 Aggregate Behavior

Table 9 summarizes the implemented actions and outcomes in the form of contingency tables.[9] As the margins show, U and L are predominantly implemented in all three treatments. In PD High they are implemented 73 % of the time each. This number increases to 79 % respectively 77 % in PD Low and reaches 96 % each in COORD. In order to test if U and L are the modal implemented actions (hypothesis 1), we conservatively consider each matching group an independent observation. If in a matching group U or L is the modal implemented action, we count it as a success. In every treatment U and L are the modal implemented actions in *all* matching groups, yielding 6 successes out of 6 tries for each treatment, generating a p-value of $p < 0.05$ (binomial tests).

[9] The two trial periods are included in the graphs but omitted from the tables and statistical analyses.

	L	R	$\sum$
U	0.51	0.22	0.73
D	0.22	0.05	0.27
$\sum$	0.73	0.27	1

PD High

	L	R	$\sum$
U	0.61	0.18	0.79
D	0.16	0.05	0.21
$\sum$	0.77	0.23	1

PD Low

	L	R	$\sum$
U	0.92	0.04	0.96
D	0.04	0.01	0.04
$\sum$	0.96	0.04	1

COORD

Note: The sums on the margins show the frequency of the implemented actions. The inner cells show the frequency of outcomes.

Table 9: Actions and Outcomes in Symmetric Treatments

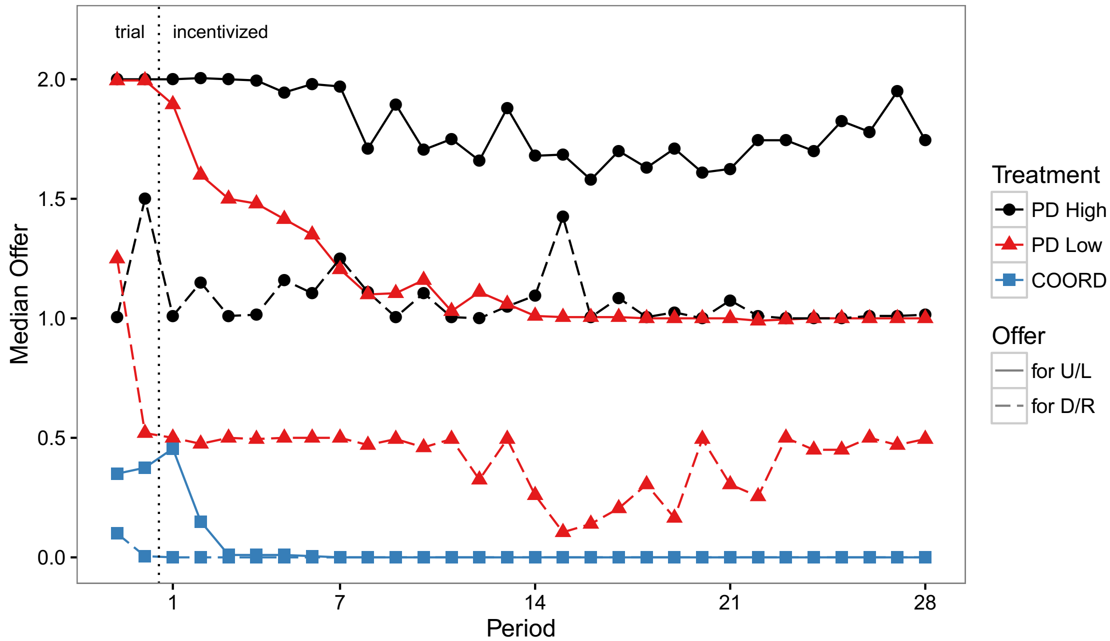
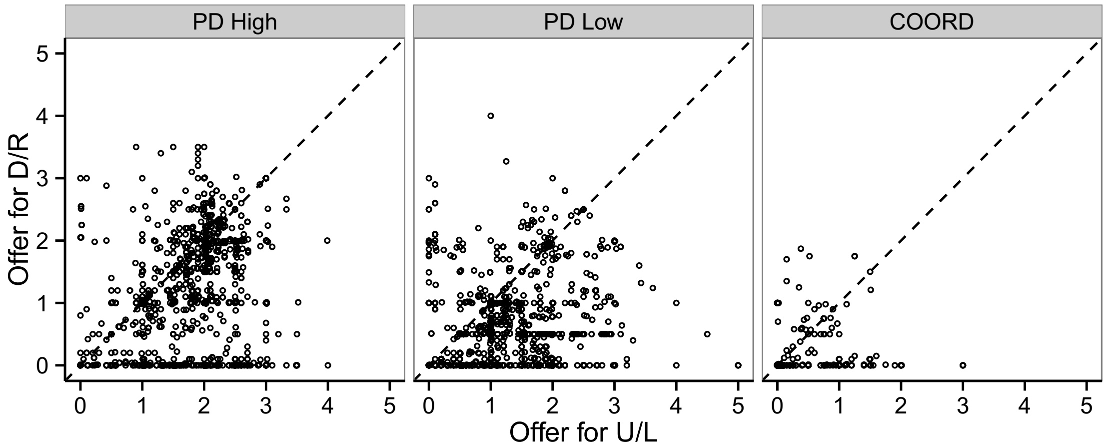


Figure 8: Median Offers to Agents

Result 1. *In treatments PD High, PD Low and COORD U and L are the modal implemented actions, lending support to hypothesis 1.*

Correspondingly, outcome UL is implemented most of the time in all treatments. The outcome share of UL is 51 % in PD High, rises to 61 % in PD Low and goes as high as 92 % in COORD. In the PD treatments there are still some deviations from UL. Subjects post offers to agents such that the agents deviate from the principals' joint profit maximum UL to UR or DL 22 % of time each. This number decreases to 18 % respectively 16 % in PD Low and finally to 4 % each in COORD. The outcome DR accounts only for 5 % of all outcomes in PD High and PD Low. In COORD outcome DR only accounts for 1 % of all cases. This is important for overall efficiency: Including the welfare of agents, efficiency levels[10] are 81 % in PD High, 78 % in PD Low and 92 % in COORD.

[10] $\text{Efficiency} = \frac{\text{achieved payoffs - minimal payoffs}}{\text{maximal payoffs - minimal payoffs}}$.



Note: The dashed line shows the 45 degree line, i.e. symmetric offers.

Figure 9: A Scatter Plot of Transfer Offers

Next we turn to the offers that subjects make to agents. First recall the predictions. Hypothesis 2 states that offers decrease from PD High to PD Low to COORD. Indeed, figure 8 shows that median offers for both U/L and D/R decrease from PD High to PD Low to COORD in all incentivized periods. The median offer for $(U/L, D/R)$ decreases from $(1.80, 1.02)$ in PD High to $(1.01, 0.49)$ in PD Low to $(0.00, 0.00)$ in COORD. These offers are significantly different between PD High and PD Low (MWU test, $p < 0.1$ and $p < 0.05$ respectively)[11]. Analogously, the offers are significantly different when comparing either PD High or PD Low with COORD ($p < 0.01$ for both auctions). All these results are in line with the theory's comparative statics.

Result 2. *Median offers decrease from PD High to PD Low to COORD, in line with hypothesis 2.*

[11]Note that in each of the 6 matching groups per treatment we observe 8 subjects making 28 repeated and incentivized decisions. In order to provide conservative estimates we consider the matching group as an independent observation, that is, we average observations on a matching group level before running the tests in this section. All tests in this article are two-sided unless explicitly stated otherwise.

Condition (AI) predicts symmetric offers with ties that are broken in favor of U and L. Only 3 % of offers are tied in both PD High and PD Low, whereas in COORD 68 % of offers are. Figure 8 already shows that median offers for U/L are remarkably close to the predicted offers of the lowest-transfer equilibria in all three treatments. But note how in all treatments the median offers for D/R are consistently lower or equal to offers for U/L and thus fall short of the symmetry prediction in PD High and PD Low. In all treatments the median offers for U/L are weakly larger than for D/R. In treatment COORD the median offer to either agent is zero, but as the graph shows in the beginning subjects still offer more for U/L than D/R. While figure 8 shows U/L and D/R independently, we plot the offers in a scatter plot in figure 9. Indeed, for PD High and PD Low the offers are predominantly on or below the 45 degree line, i.e. offers are higher for U/L than for D/R. In COORD though, nearly all actions are at or extremely close to zero. In other words, the symmetry prediction seems to work for COORD but not so much for the PD treatments. The difference between offers for D/R and U/L is significant (Wilcoxon Signed-Rank, $p < 0.05$ for PD High and PD Low, and, surprisingly, also $p < 0.1$ for COORD) for all treatments. Thus, we have to reject the prediction from hypothesis 3. We will investigate this phenomenon more closely in section 3.3.4 where we study quantal response equilibria in our games.

Result 3. *Median offers for U/L are larger than for D/R. Offers for U/L follow theoretical predictions but offers for D/R fall short. We reject hypothesis 3.*

Finally, figure 8 indicates that offers decrease weakly over time in all treatments. We test if offers decrease from the first to the second half of the experiment. For PD High the difference is significant for offers for U/L (Wilcoxon Signed-Rank, $p < 0.1$) but not for for offers for D/R ($p > 0.1$). For PD Low and COORD the difference is significant for both U/L and D/R ($p < 0.05$). While offers decrease a little outcomes are remarkably stable. Table 23 in appendix C shows that the distribution of outcomes does not change much from the first to the second half of the experiment, while offers decrease somewhat.

3.3.3 Individual Learning

Given the complex nature of equilibrium, we were surprised to see aggregate behavior approaching equilibrium play so closely. Of course, the failure of the symmetry prediction tells us that subjects do not really acquire equilibrium reasoning. Rather, we must look for a combination of initial heuristics and a learning process if we want to understand how behavior develops.

Regarding initial heuristics let us inspect figure 8 again. In all three treatments subjects start out with considerably higher offers for U/L than for D/R. With a little introspection this is perhaps not very surprising. In all three games, implementing U is more important for A principals than implementing D. For example, in PD High, the gain from U is at least 4 while the gain from D is at most 2. So, focusing on the more important auction appears natural. (Indeed, we will revisit this issue—that making a mistake on U is much worse than making a mistake on D—when we estimate quantal response equilibria.) Figure 8 also indicates a potential negative time trend for some offers while in the second half offers seem to stabilize. Indeed, in the first half of the experiment there is a significantly negative time trend for offers for U/L in all three symmetric treatments as well as for offers for D/R in COORD. In the second half, though, there is no significant time trend in any of the treatments. Tables 27 and 28 in appendix C summarize the relevant regression results.

With regard to the learning process that unfolds in the first half of the experiment, we take our cue from Selten's learning direction theory, that is, the idea that sub-

Change	-	NA	+	∑
+	81	4	95	180
0	166	11	58	235
-	198	1	10	209
∑	445	16	163	624
	LDT			

U/L PD High

Change	-	NA	+	∑
+	14	106	43	163
0	58	170	70	298
-	89	55	19	163
∑	161	331	132	624
	LDT			

D/R PD High

Change	-	NA	+	∑
+	61	3	88	152
0	169	10	47	226
-	233	2	11	246
∑	463	15	146	624
	LDT			

U/L PD Low

Change	-	NA	+	∑
+	21	67	20	108
0	63	273	54	390
-	62	63	1	126
∑	146	403	75	624
	LDT			

D/R PD Low

Change	-	NA	+	∑
+	13	7	12	32
0	116	320	22	458
-	130	4	0	134
∑	259	331	34	624
	LDT			

U/L COORD

Change	-	NA	+	∑
+	6	6	3	15
0	10	533	25	568
-	26	15	0	41
∑	42	554	28	624
	LDT			

D/R COORD

Note: On the x-axis LDT shows learning direction theory's predictions. The theory either prescribes a lower offer (-), a higher offer (+) or doesn't yield a prediction (NA). On the y-axis Change shows whether a subject increased (+), decreased (-) or didn't change her offer (0) in the following period. Each cell counts the number of cases. Cells are colored green (dark gray when printed in black and white) if the theory correctly predicted the change, yellow (light gray when printed in black and white) if not, and white if it didn't yield a prediction. Here, we count zeros as successes. Note that we do not use the data from the non-incentivized trial periods.

Table 10: Learning Direction Theory in Periods 1-14

jects move towards myopic better replies (see for a similar application to learning in auctions Selten et al. (2005)). In the context of our intertwined auctions, the theory simply predicts that a subject will (i) lower her offer if she either won an auction and could have won with a lower offer or if she had offered more than she subsequently gained; and (ii) increase her offer if she lost an auction but could have made a profit by winning it. This leaves two cases. First, a subject may have profitably won an auction by marginally (that is, optimally) offering more than her opponent. Second, a subject may have lost an auction and could not have profitably deviated because her opponent posted an offer above her reservation value. Any subsequent offer that is lower or equal (or potentially a bit higher) is then a myopic best reply. In these two cases the theory is silent on the direction of an adjustment.

In order to check for directional learning in this spirit we conduct a simple counting exercise. We count how often subjects changed their offer from one period to the next, and how often their behavior is in line with the predictions of learning direction theory. Table 10 shows panels with contingency tables for the first half for all treatments, because this is when we observe a significant time trend in our regressions. We consider a weak version of learning direction theory and only consider change *against* the prescribed direction as a violation of learning direction theory, counting zero change as a weak success. Cells count the number of cases and are colored green[12] if learning direction theory successfully predicted the change and yellow[13] if not. The theory turns out to be a good predictor for subjects' behavior if one compares how often subjects move into the predicted rather than the opposite direction. Subjects often stay, but mostly go in the predicted direction. They very seldom go into the opposite direction of what learning direction theory would predict, and this holds for

[12] Dark gray when printed in black and white.

[13] Light gray when printed in black and white.

all treatments. In PD High 86 % of all adjustments follow the predictions of learning direction theory, which increases to 89 % in PD Low and 95 % in COORD. Also, note how learning direction theory typically gives clear predictions in PD High and PD Low for offers on U/L. Only 16 respectively 15 observations out of 624 do not come with a prediction. On the other hand, for offers on D/R this number increases to 331 respectively 403 out of 624 observations. This pattern stems from subjects tending to lose the auction for the D/R action so decisively that they could not have gained from either increasing or decreasing their offer. This finding suggests an intuitive explanation for why offers for D/R remain lower than offers for U/L. As subjects could not have gained anyway, there is no force driving them to increase their offers.[14]

3.3.4 Quantal Response Equilibrium

The theoretical predictions for our three symmetric games work well for outcomes but offers on D/R are well below predictions in the two PD treatments. Remember that the lowest-transfer equilibrium offers for $(U/L, D/R)$ in PD High, PD Low and COORD are offers of $(2, 2)$, $(1, 1)$ and $(0, 0)$ respectively. The median offers for U/L are close to the equilibrium predictions but offers for D/R are significantly smaller than the former. Only in COORD the equilibrium predictions for offers fit the data precisely. As discussed above, this appears to stem from some simple heuristic reasoning that encourages principals to offer more in the more important auction—a pattern that is maintained also in the presence of learning (simply because there is not much pressure to increase the offer in an auction one is supposed to lose anyway). In other words, deviations appear to occur when they are not so costly. In order to organize

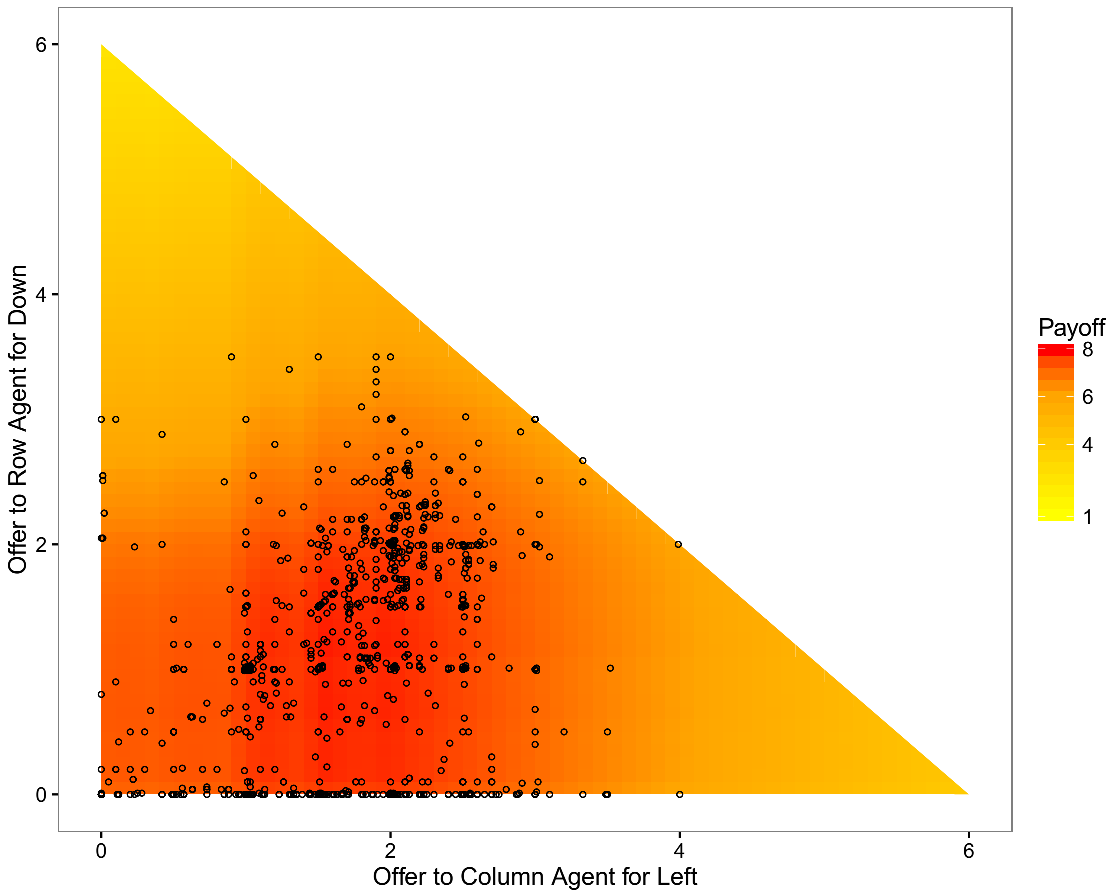
[14] In the second half of the experiment learning direction theory loses some of its bite as subjects tend to stay more than before. Results are provided in table 26 in appendix C.

this pattern—that subjects make more errors when they are less costly—estimating a quantal response equilibrium model appears the natural way forward.

Consider a principal A in PD High who expects principal B to choose the symmetric equilibrium offer of $(2, 2)$. If A also offers 2 to both agents, this will implement outcome UL. Principal A ends up only paying the column agent, losing the auction for the row agent. The offer of 2 to the row agent only weakly dominates offers below 2, so any potential downward deviations are costless in equilibrium. On the other hand, deviations from the offer of 2 to the column agent are strictly costly: offering too little means losing the auction for the column agent and offering too much means paying too much. As pointed out before, these considerations are suggestive and models that capture the costs of errors should, hence, organize the data and explain the gap in offers for U/L and D/R. The same reasoning holds for PD Low. Moreover, it also provides an explanation for why offers are so close to equilibrium predictions in COORD as in this treatment *all* deviations from the equilibrium offer of 0 are costly. Still, we fit the QRE to all three treatments.

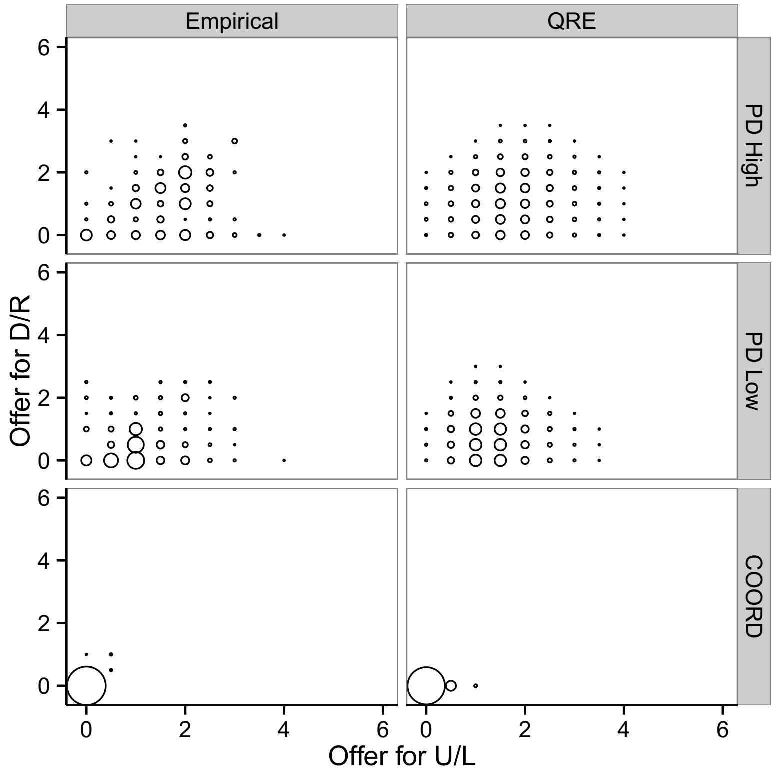
In figure 10 we draw the expected payoff heatmap of a principal A playing against the empirical distribution of offers in PD High. It shows an area of approximately maximal payoffs, the darkest part of the figure, around the perpendicular on an offer for (L, D) of $(2, 0)$, bounded above by the equilibrium offer of $(2, 2)$. This corridor happens to contain most actual offers indicating that subjects play noisy best replies against the empirical distribution.

We formalize this intuition in a QRE framework following McKelvey and Palfrey (1995). We do this in order to show that the above intuition can be captured in a consistent equilibrium framework. As mentioned before there is evidence that offers change from the first to the second half of the experiment. Figure 8 and correspondingly table 23 in appendix C) show a time trend in the first half of the experiment



Note: Each circle represents one observed offer. Given these offers we calculate the expected payoff of every feasible combination of offers and draw the expected payoffs heatmap.

Figure 10: Expected Payoffs for Principal A in PD High



Note: The area of a bubble corresponds to the probability of the corresponding combination of offers. Very small bubbles may not be displayed.

Figure 11: Empirical and Best-Fit QRE Distribution of Offers

which is explained by directional learning. Regressions show a significant time trend in some of the offers in the first half of the experiment which dies down in the second half (see tables 27 and 28 in appendix C) as behavior settles down and converges to equilibrium. Consequently, in what follows we consider only periods 15 to 28.[15] More details on the setup of the QRE as well as some robustness checks are documented in appendix C).

The best fits are obtained for a precision parameter $\lambda = 1.91$ for PD High, $\lambda = 2.27$ for PD Low and $\lambda = 5.23$ for COORD. The QRE models fit the data well. Figure 11 juxtaposes the empirical distribution of offers with the predicted probabilities in the

[15]See Goeree et al. (2002) for a similar procedure.

best-fitting QRE. The area of a bubble corresponds to the probability of the corresponding combination of offers. Note how closely the distributions resemble each other. Empirically, there are a lot of offers different from $(2, 2)$ in PD High with more distribution mass on low offers for D/R than on low offers for U/L. Similarly, in the QRE with the best fit there is a lot of probability mass on low offers for D/R but less so for offers for U/L. These results are in line with our earlier intuition. Errors for offering too little for D/R are not very costly given the other subject's offer but conversely for U/L they are. A similar pattern can be observed in PD Low but now probability mass is shifted to lower offers in both dimensions, which is in line with theoretical predictions. Finally, for COORD nearly all observations are at or extremely close to $(0, 0)$ which is also reflected in the corresponding QRE.

Result 4. *Subjects' overall choice patterns are consistent with subjects making errors in a quantal response equilibrium framework.*

In this section we have shown how quantal response equilibrium can make sense of the behavioral patterns in the data. But, obviously, we can explain a lot of behavior when fitting a model to data ex post. Consequently, we designed an asymmetric follow-up treatment to test some predictions of QRE out of sample.

3.4 Asymmetric Games

All games so far were symmetric. In this section we study two asymmetric settings, one to test ex ante predictions of quantal response equilibrium and another one to test if the theory has bite in a more intricate setting. The first asymmetric treatment features a gross payoff matrix corresponding to a simple dominance solvable game that we will use to put QRE to a test. In the symmetric PD treatments we observe an asymmetry within a principal's offers as principals typically offer more for L than

	L	R
U	$5,0$	$0,1$
D	$0,1$	$0,2$

DOM

	L	R
U	$5,2$	$0,0$
D	$0,0$	$2,4$

BoS

Table 11: Asymmetric Principals' Payoff Matrices

for D. In the first asymmetric treatment we predict instead an asymmetry of offers *between* the two types of principals. We refer to this treatment as DOM. The second, more intricate, asymmetric treatment has a gross payoff matrix corresponding to a battle of the sexes. We refer to this treatment as BoS. Both payoff matrices are shown in table 11.

Consider treatment DOM first. In this treatment principal A offers transfers for actions U and L while principal B offers transfers for D and R. Outcome UL is the most efficient so theory predicts it to arise in equilibrium. The corresponding offers that support this outcome are all equal to 1. Implementing R instead of L or D instead of R translates to a net increase in gross payoffs for principal B of 1, so principal A should offer 1 both for U and L to discourage principal B from overbidding her. In equilibrium, principal B matches these offers and the ties are broken in favor of U and L.

Let us formally derive the equilibrium for this game. The unique efficient outcome in this game is UL, hence it must result in any pure-strategy Nash equilibrium. By *IC* we know that

$$5 \geq t_R^A(U) + t_C^A(L)$$

$$0 \geq 1 - t_R^B(D)$$

$$0 \geq 1 - t_C^B(R)$$

Obviously, the offers for D and R have to be at least 1 each and thus by **(AI)** the offers for U and L have to be at least 1 each but may not sum up to more than 5. The set of pure-strategy equilibria with UL as an outcome is

$$(t_R^A(U) = t_R^B(D), t_C^A(L) = t_C^B(R)) \in \{(a, b) \mid a + b \leq 5 \wedge a, b \geq 1\}$$

There exists a unique equilibrium with the lowest total transfer from the principals:

$$t_R^A(U) = t_C^A(L) = t_R^B(D) = t_C^B(R) = 1$$

Let us now turn to treatment BoS. The unique efficient outcome in this game is UL, hence it must result in any pure-strategy Nash equilibrium. By *IC* we know that

$$2 + t_R^A(U) + t_C^A(L) \geq max\{t_R^A(U), t_C^A(L), 4\}$$

$$5 \geq max\{t_R^B(D), t_C^B(R), 2 + t_R^B(D) + t_C^B(R)\}$$

By the **(AI)** condition, in equilibrium, $t_R^A(U) = t_R^B(D)$ and $t_C^A(L) = t_C^B(R)$. Thus, the conditions above simplify to

$$5 \geq 2 + t_R^B(D) + t_C^B(L)$$

$$4 \leq 2 + t_R^A(U) + t_C^A(R)$$

and the set of pure-strategy equilibria with UL as an outcome is

$$(t_R^A(U) = t_R^B(D), t_C^A(L) = t_C^B(R)) \in \{(a, b) \mid a + b \in [2, 3]\}$$

There exists a unique *set* of equilibria with the lowest total transfer from the principals:

$$t_R^A(U) + t_C^A(L) = t_R^B(D) + t_C^B(R) = 2$$

3.4.1 Experimental Implementation and Hypotheses

The experimental procedures were nearly the same as before. The main difference is that we now also employed different instructions such that the principals would see the game from their own perspective. Subjects are assigned the role of principal A or principal B at the beginning of the game and they keep their role throughout the experiment. As before, subjects are randomly assigned to matching groups of eight (with four principals A and B each) and are rematched every round, playing again for 2 trial and 28 incentivized rounds. In these games, principal A can incentivize the column agent to choose L and the row agent to choose U. Conversely, principal B can incentivize the column agent to choose R and the row agent to choose D. Subjects in both treatments were endowed with a budget of 5. In DOM we ran two sessions of 24 subjects each. In BoS we ran four sessions. Three sessions were run with 24 subjects, but in a fourth session too few subjects showed up and the session was run with 20 subjects, with one matching group of eight and two matching groups of six. In our analysis we drop the entire session, but all results are robust to including these observations.

Now consider what QRE predicts in treatment DOM. Principal A stands to lose a lot from losing either U or L. Therefore, as precision increases, QRE predicts her offers to be very close to the theoretical prediction of 1, so virtually the same prediction as standard equilibrium analysis. Principal B on the other hand can expect to lose both actions D and R so she is indifferent between offering 1 or less for either action. Therefore, QRE predicts principal B on average to bid less than principal A on both actions. This predicted asymmetry between principals' offers is markedly different

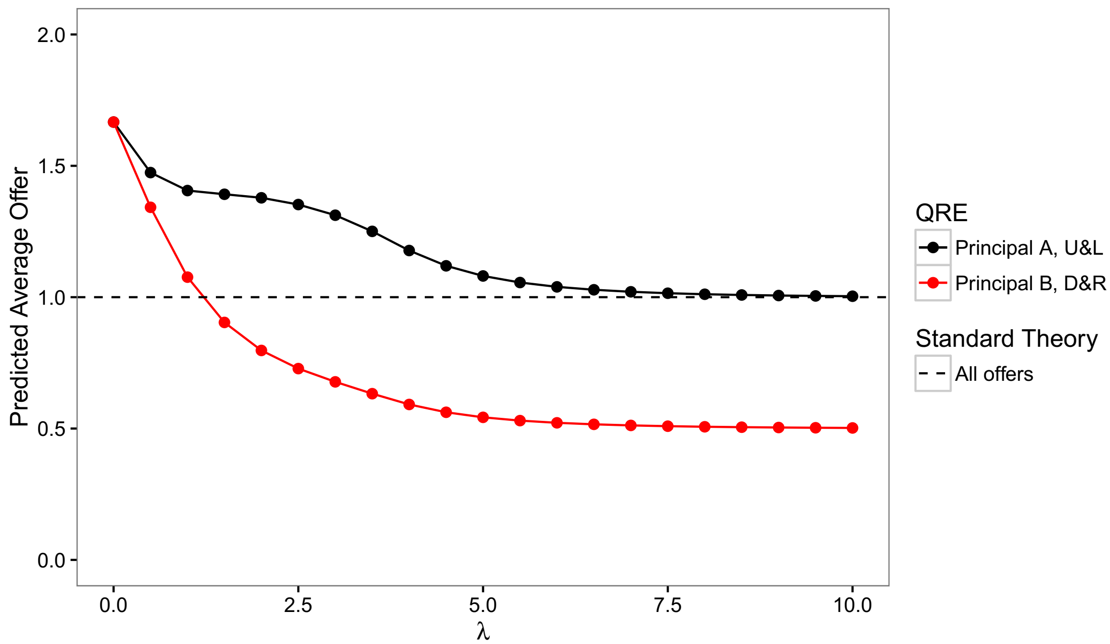


Figure 12: Predicted Average QRE Offers

from QRE predictions in the symmetric games where we expect an asymmetry within a principals' offers. Figure 12 shows how QRE's predicted offers for both principals change as the precision parameter λ increases. Note that, on average, principal A's offers for U and L should be the same as well as principal B's offers for D and R. At $\lambda = 0$ both principals perfectly randomize between all available offers and the average offer is about 1.7. As λ increases average offers decrease, though initially more rapidly for principal A than for principal B. As $\lambda \to \infty$ principal A's offers converge to the standard theory prediction of 1 for both U and L. Principal B on the other hand expects to lose against principal A's offers and is indifferent between all offers below 1. Hence, as $\lambda \to \infty$ principal B's average offers converge to 0.5 for both D and R. From this graph it is easy to see that for any $\lambda > 0$ QRE predicts principal A's offers to be larger than principal B's offers while standard theory predicts all offers to be the same. Furthermore, note that since the offers for U and L are bigger than the offers for D and R, both QRE and standard theory predict that U and L should typically be implemented.

Hypothesis 4. *U and L are the modal implemented actions.*

Hypothesis 5. *All offers tie. (Standard theory's null hypothesis)*

Hypothesis 6. *Offers for D respectively R are smaller than offers for U respectively L. (QRE's alternative hypothesis)*

For the BoS game the predictions are not as crisp. Before, we had a unique lowest-transfer equilibrium and a point prediction for the individual offers. Now we have a point prediction for the total offer but not for the individual offers. Any combination of offers that sum up to 2 can support a lowest-transfer equilibrium. Consequently, the BoS game exhibits two added layers of complexity compared to the symmetric games. First, there is now more conflict from the outset. Second, subjects also have to coordinate on how to allocate the sum of offers. After the strong performance of Prat and Rustichini (2003)'s theory for symmetric games, the BoS game serves very much a stress test. The predictions are

Hypothesis 7. *U and L are the modal implemented actions.*

and

Hypothesis 8. *Offers to both agents sum to 2 for both principals.*

3.4.2 Results

As in the symmetric treatments, Prat and Rustichini's (2003) theory does a remarkable job predicting the implemented actions and resulting outcomes as table 12 shows. In treatment DOM, U and L are the implemented actions 81 % respectively 82 % of

	L	R	$\sum$
U	0.72	0.10	0.82
D	0.08	0.09	0.18
$\sum$	0.81	0.19	1

DOM

	L	R	$\sum$
U	0.59	0.02	0.61
D	0.02	0.37	0.39
$\sum$	0.61	0.39	1

BoS

Note: The sums on the margins show the frequency of the implemented actions. The inner cells show the frequency of outcomes.

Table 12: Actions and Outcomes in Asymmetric Treatments

the time. In all 6 matching groups U and L are the modal implemented actions, providing evidence for hypothesis 4 (two-sided binomial tests, $p < 0.05$). In BoS, U and L are the implemented actions 61 % of the time each. In 7 out of 9 matching groups U and L are the modal implemented actions, providing evidence for hypothesis 7 (one-sided binomial tests, $p < 0.1$)[16].

Result 5. *In both DOM and BoS U and L are the modal implemented actions, supporting hypotheses 4 and 7.*

Corresponding to the implemented actions, in DOM 72 % of outcomes are UL. Only 18 % of outcomes are either DL or UR and only 9 % of the time principal B can implement her most preferred outcome DR. As even minor deviations from UL are very costly from an efficiency standpoint subjects achieve an efficiency level of only 74 %. Turning to BoS, 59 % of outcomes are UL with DR coming up 37 % of the time and remarkably little miscoordination on either UR or DL. These results are interesting from an efficiency point of view as subjects successfully manage to avoid miscoordination. Subjects achieve a remarkable efficiency level of 90 %. The distribution of outcomes is also quite stable in either treatment and does not change much over time as tables 24 and 25 in appendix C show.

[16]Note that we construct our tests very conservatively by dropping one session with too few subjects, aggregating on the matching group as the independent observation and making no distributional assumptions. If one includes the omitted session the success ratio for hypothesis 7 increases to 10 out of 12 (two-sided binomial tests, $p < 0.05$).

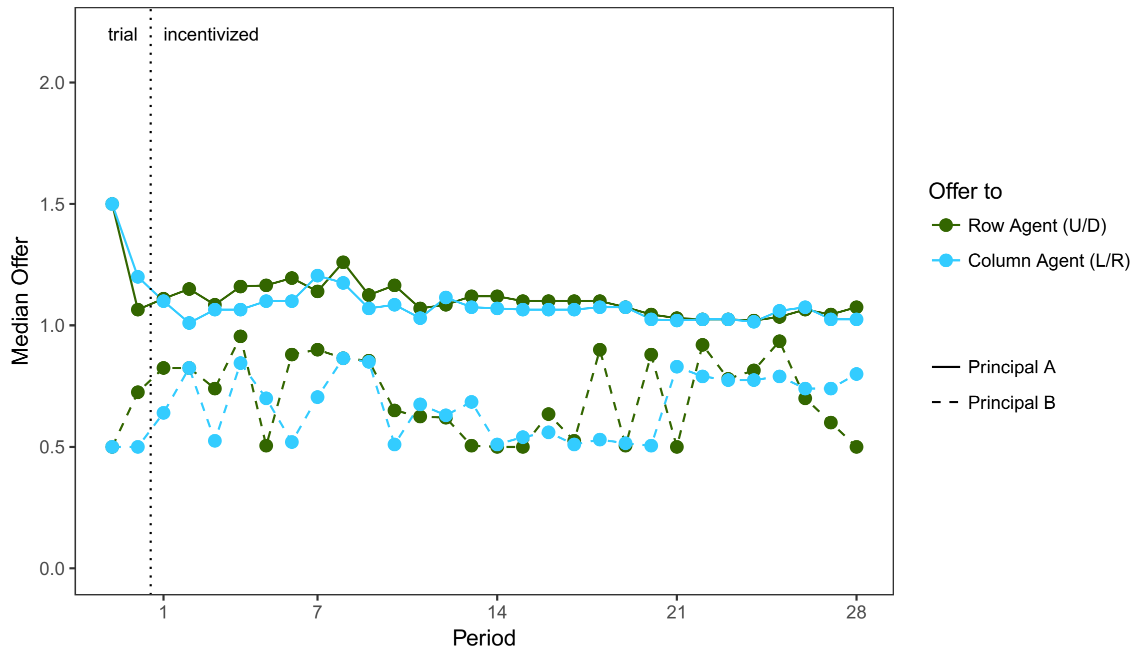


Figure 13: Median Offers in DOM

Let us now turn to the offers. In DOM standard theory would predict all offers to tie. Figure 13 shows that this is not the case. Principal A's median offers for (U, L) are $(1.05, 1.10)$ while principal B's median offers for (D, R) are $(0.655, 0.735)$. So, on the one hand, principal A's offers for U and L are not significantly different from each other as are principal B's offers for D and R (Wilcoxon Rank-sum tests, $p > 0.1$). But on the other hand, principal A's offers for U and L are each significantly larger than principal B's offers for D and R (Wilcoxon Rank-sum tests, $p < 0.05$). Hence, we can reject standard theory's null hypothesis 5 but we cannot reject QRE's alternative hypothesis 6. Examining the offers more closely, principals show a taste for symmetry. Principal A offers symmetric transfers to both agents 73 % of the time while principal B offers a symmetric transfer 50 % of the time. Indeed, in a QRE framework it makes sense for principal A to be more interested in offering a symmetric transfer than principal B. Principal B typically loses against principal A,

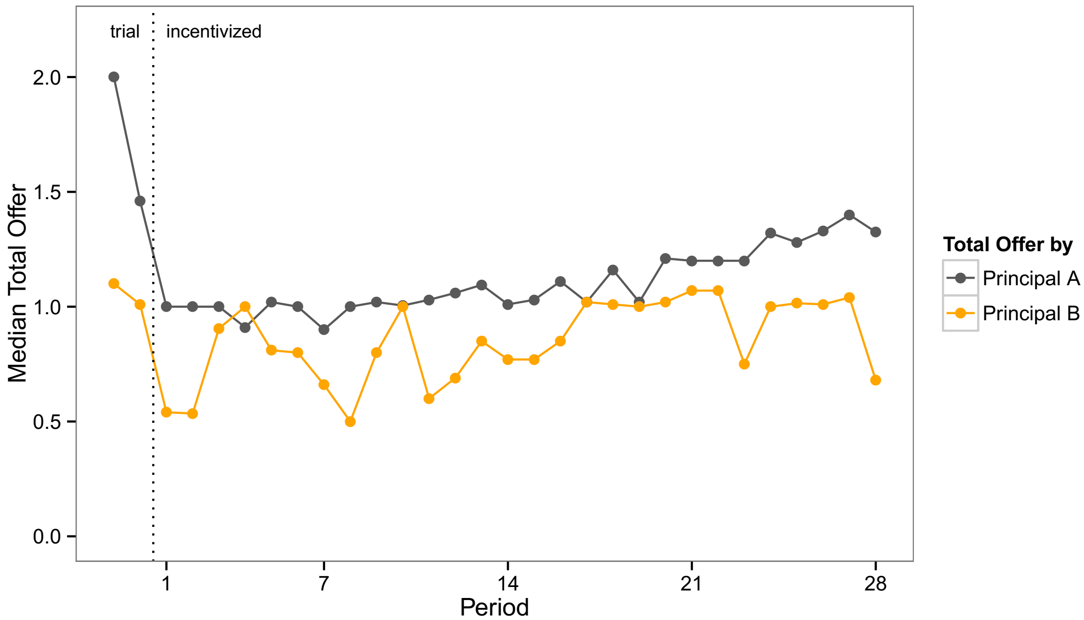


Figure 14: Median Total Offers in BoS

so, in terms of payoffs, it makes little difference if she is to lose with an asymmetric or a symmetric bid. Principal A on the other hand does not want to be overbid, but offering more than necessary is costly as well. Hence, we can intuitively expect more symmetric offers by principal A than by principal B. This pattern can be seen in Figure 24 in appendix C, which shows a number of QREs for increasing λs.

Result 6. *In DOM offers for U and L are larger than offers for D and R. Hence, we have to reject standard theory's null hypothesis 5 in favor of QRE's alternative hypothesis 6.*

When looking at the offers in BoS let us first check how subjects deal with the issue of coordination. Even more often than in treatment DOM, subjects typically choose the same offers for the two actions. 81 % of the time principals choose to offer the same amount to both agents. Choosing symmetric offers appears to be a focal point on which subjects coordinate. Intuitively, one can see the appeal of such offers.

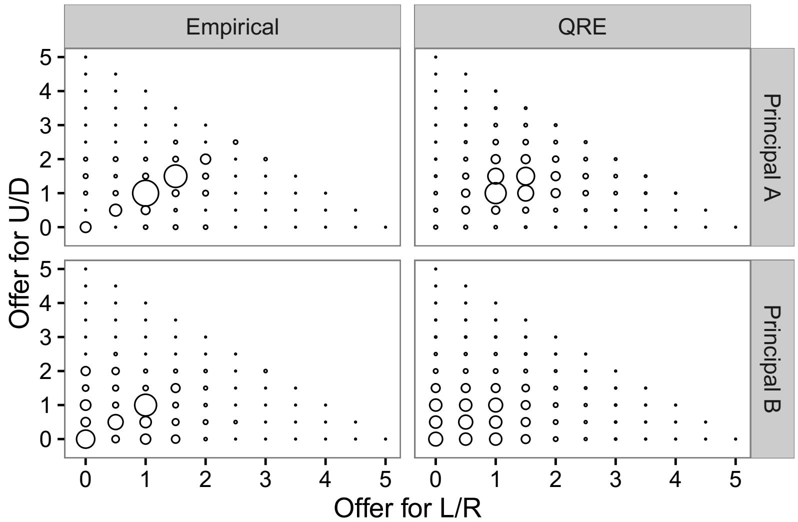


Figure 15: Empirical and Best-Fit QRE Distributions for DOM

Given that other subjects make symmetric offers it is best to place a symmetric offer as well. Either one wins with both offers and ends up with one's preferred outcome, or one loses and ends up with the opponent's preferred outcome. Either outcome is preferable to miscoordination on UR or DL though, which would be the potential result of posting a non-symmetric offer. As our predictions concern total offers and most offers are symmetric anyway we can focus on total offers. Figure 14 shows the evolution of total offers over the course of the experiment. Total offers to agents seem to increase somewhat but remain below the equilibrium prediction of 2. Furthermore, principal A's median total offer is 1.08 which is significantly higher than principal B's median total offer of 0.82 (Wilcoxon Rank-sum, $p < 0.1$).[17]

Result 7. *In BoS median total offers are below equilibrium predictions and principal A offers more than principal B. Therefore we have to reject hypothesis 8.*

Subjects do not seem to engage much in learning in either asymmetric treatment

[17] As before we average observations on a matching group level before running the tests in this section.

as offers do not significantly increase from the first half to the second half of the experiment (Wilcoxon Rank-sum tests, $p > 0.1$). Medians are also stable over time as tables 24 and 25 in appendix C show. Similarly, a linear regression of periods on offers shows no significant time trend.[18] We fit QRE on the data of all 28 periods.[19] In DOM the best-fit QRE has a precision parameter of $\lambda = 3.00$. Figure 15 shows the offers that we observe in the experiment and the offers of the best-fit QRE. Note how closely the distributions resemble each other, even though we observe slightly more symmetric offers in the data than we would expect from QRE. The best fit in the BoS treatment is for $\lambda = 1.62$.[20] Figure 16 summarizes the results. The QRE does not fit as well as in the case of the symmetric treatments but still captures the essential features of the data. As mentioned earlier, subjects predominantly use symmetric offers, but there is no force in the QRE that would ensure such symmetry. Consequently, in the corresponding QRE there is too much probability mass on asymmetric offers. As principal A tends to offer more than principal B, principal B randomizes on low symmetric as well as asymmetric offers. Still, principal A is predicted to offer more than principal B, which is reflected in the data.

Result 8. *In DOM quantal response equilibrium organizes subjects' offers very well. In BoS it captures the central result of principal A offering more than principal B, but fails to yield symmetric offers.*

[18] Regression results are reported in tables 29 and 30 in appendix C.

[19] In the asymmetric games there are two types of principals instead of just one so we adapt our estimation strategy accordingly. Details and robustness checks are documented in appendix C

[20] As above, we also calculate a number of QREs for an increasing λ in order to understand how the process converges. Results are provided in figure 25 in appendix C.

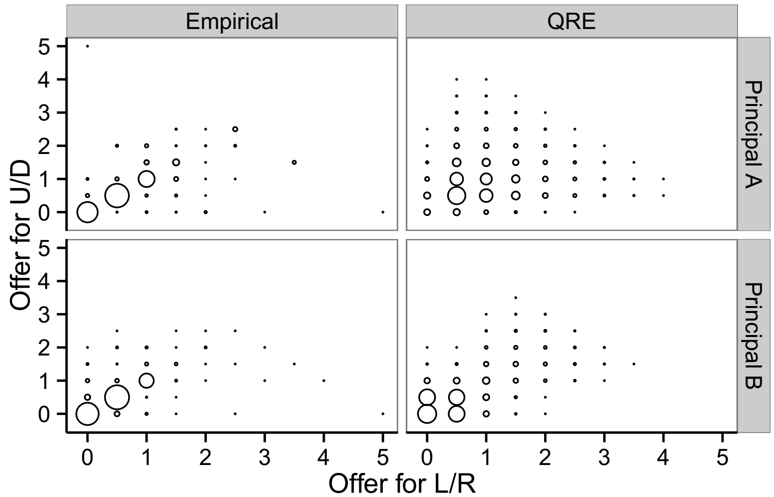


Figure 16: Empirical and Best-Fit QRE Distribution of Offers for BoS

3.5 Discussion

This paper presents an experimental test of Prat and Rustichini (2003). Examining situations where multiple principals can influence multiple agents by promising monetary transfers, their model captures an important class of real-world strategic interactions. Yet, so far the model has remained empirically untested.[21] Despite its non-trivial equilibrium logic, the model predicts behavior extremely well. In the symmetric treatments and the asymmetric DOM treatment the theory successfully predicts which actions are implemented about 80 % of the time. In the more demanding asymmetric BoS treatment the theory succeeds more than 60 % of the time.

While our results for implemented actions provide strong support for the theory, subjects' offers require closer scrutiny. Comparative static predictions for the symmetric games that we implemented hold. Also, behavior quickly settles down as subjects learn to play the game. The observed learning pattern is well organized by

[21]Rassenti et al. (1982) study combinatorial auctions of objects when bidders have preferences over bundles of objects. This can be seen as a special case of Prat and Rustichini (2003) where payoffs depend only on the auctions that one wins. In GPTAs payoffs would generally also depend on the auctions that one loses.

Selten's learning direction theory. While the naive prediction of all probability mass being concentrated on one point fails, deviations arise in a clear pattern. On the one hand, offers for actions that are implemented (such that one has to pay for them) are very close to the theory's predictions. On the other hand, offers for actions that are typically not implemented (and thus are without immediate payoff consequences) are lower than predicted. Instead of making offers for the latter such that their opponent becomes indifferent subjects bid too little violating the agent indifference condition of Prat and Rustichini's (2003) theory. To explain the deviations we apply a quantal response equilibrium framework to our data. QRE can make sense of the asymmetry in offers that we observe and allows us to successfully predict behavior out of sample in a follow-up treatment. In treatment DOM we test for an asymmetry between two types of principals' offers as a central *ex ante* prediction of QRE. We can reject symmetry of offers in favor of QRE's predictions. Integrating QRE into GPTAs is natural due to the inherent asymmetry in local incentives at the equilibrium. Finally, it should be noted that, on the level of implemented outcomes and actions, QREs are, for the usual range of the precision parameter, remarkably close to Nash.

One might be tempted to compare our data to behavior in simple 2×2 games played without intermediaries. We do not offer such a comparison as it is a red herring. The whole idea of GPTAs is to look at situations where principals cannot play the base game (just as lobbyists cannot vote in parliament). Also, principals do not simply hire an agent to play the game for them. Rather, they influence other people who play a game whose outcome has externalities—on the principals. Moreover, the game that these others, the agents in the GPTA terminology, play is in our case a 2×2 game where all players receive zero payoffs in all cells, that is, the base game is neither a prisoners' dilemma, nor a battle of the sexes game.

Our study should be seen as a first step towards more research into GPTAs. We only test relatively simple 2×2 interactions (2 principals, 2 (computerized) agents with 2 strategies each). Including human agents appears to be one of the more desirable extensions for future work. Of course, this route is likely to require the consideration of social preferences which are absent from the base model. Other extensions would examine more complicated types of interaction. How far can the theory be pushed in explaining who gets their way in such games of economic influence? Voting games might appear a particular attractive avenue given the model's natural applications to lobbying. Our results suggest that, just as theory predicts, principals tend to incentivize agents such that the socially efficient outcome results most of the time. It strikes us as remarkable that this happens even in situations of strong conflict such as in our asymmetric treatments.

References

Aidt, Toke S. and Uk Hwang (2008) "On the Internalization of Cross National Externalities Through Political Markets: The Case of Labour Standards," *Journal of Institutional and Theoretical Economics*, Vol. 164, pp. 509–533, DOI: `http://dx.doi.org/10.1628/093245608785363380`.

Apesteguia, Jose, Steffen Huck, and Jörg Oechssler (2007) "Imitation—theory and experimental evidence," *Journal of Economic Theory*, Vol. 136, pp. 217–235, URL: `http://www.sciencedirect.com/science/article/pii/S0022053106001268`, DOI: `http://dx.doi.org/10.1016/j.jet.2006.07.006`.

Axelrod, Robert and William D. Hamilton (1981) "The evolution of cooperation," *Science*, Vol. 211, pp. 1390–1396, URL: `http://science.sciencemag.org/content/`

211/4489/1390.abstracthttp://www.jstor.org/stable/1685895, DOI: http://dx.doi.org/10.1126/science.7466396.

Bellemare, Charles and Sabine Kröger (2007) "On representative social capital," *European Economic Review*, Vol. 51, pp. 183–202, DOI: http://dx.doi.org/10.1016/j.euroecorev.2006.03.006.

Berg, Joyce, John Dickhaut, and Kevin McCabe (1995) "Trust, Reciprocity, and Social History," URL: http://linkinghub.elsevier.com/retrieve/pii/S0899825685710275, DOI: http://dx.doi.org/10.1006/game.1995.1027.

Bikhchandani, Sushil (1999) "Auctions of Heterogeneous Objects," *Games and Economic Behavior*, Vol. 26, pp. 193–220, DOI: http://dx.doi.org/10.1006/game.1998.0659.

Bordignon, Massimo, Luca Colombo, and Umberto Galmarini (2008) "Fiscal federalism and lobbying," *Journal of Public Economics*, Vol. 92, pp. 2288–2301, DOI: http://dx.doi.org/10.1016/j.jpubeco.2008.05.003.

Bowles, Samuel and Herbert Gintis (2002) "Homo reciprocans," *Nature*, Vol. 415, pp. 125–128, URL: http://dx.doi.org/10.1038/415125a, DOI: http://dx.doi.org/10.1038/415125a.

Boylan, Richard T. and Mahmoud A. El-Gamal (1993) "Fictitious Play: A Statistical Study of Multiple Economic Experiments," *Games and Economic Behavior*, Vol. 5, pp. 205–222, DOI: http://dx.doi.org/06/game.1993.1011.

Cheung, Yin-Wong and Daniel Friedman (1997) "Individual Learning in Normal Form Games: Some Laboratory Results," *Games and Economic Behavior*, Vol. 19, pp. 46–76, DOI: http://dx.doi.org/10.1006/game.1997.0544.

Cox, James C., Daniel Friedman, and Steven Gjerstad (2007) "A tractable model of reciprocity and fairness," *Games and Economic Behavior*, Vol. 59, pp. 17–45, DOI: http://dx.doi.org/10.1016/j.geb.2006.05.001.

Croson, Rachel and Uri Gneezy (2009) "Gender Differences in Preferences," *Journal of Economic Literature*, Vol. 47, pp. 448–474, URL: http://pubs.aeaweb.org/doi/10.1257/jel.47.2.448, DOI: http://dx.doi.org/10.1257/jel.47.2.448.

Dal Bó, Pedro and Guillaume R Fréchette (2011) "The Evolution of Cooperation in Infinitely Repeated Games: Experimental Evidence," *American Economic Review*, Vol. 101, pp. 411–429, DOI: http://dx.doi.org/10.1257/aer.101.1.411.

Dittrich, Marcus (2015) "Gender differences in trust and reciprocity: evidence from a large-scale experiment with heterogeneous subjects," *Applied Economics*, Vol. 47, pp. 3825–3838, URL: http://dx.doi.org/10.1080/00036846.2015.1019036, DOI: http://dx.doi.org/10.1080/00036846.2015.1019036.

Dohmen, Thomas J, Armin Falk, David Huffman, and Uwe Sunde (2009) "Homo Reciprocans: Survey Evidence on Prevalence, Behaviour and Success," *Economic Journal*, Vol. 119, pp. 592–612, URL: http://ssrn.com/abstract=936216, DOI: http://dx.doi.org/10.1111/j.1468-0297.2008.02242.x.

Dufwenberg, Martin and Georg Kirchsteiger (2004) "A theory of sequential reciprocity," *Games and Economic Behavior*, Vol. 47, pp. 268–298, DOI: http://dx.doi.org/10.1016/j.geb.2003.06.003.

Ensthaler, Ludwig, Steffen Huck, and Johannes Leutgeb (2017) "Games played through agents in the laboratory: A test of Prat & Rustichini's model."

Falk, Armin (2007) "Gift exchange in the field," *Econometrica*, Vol. 75, pp. 1501–1511, DOI: http://dx.doi.org/10.1111/j.1468-0262.2007.00800.x.

Falk, Armin and Urs Fischbacher (2006) "A theory of reciprocity," *Games and Economic Behavior*, Vol. 54, pp. 293–315, DOI: http://dx.doi.org/10.1016/j.geb.2005.03.001.

Fehr, Ernst, Urs Fischbacher, and Simon Gächter (2002) "Strong Reciprocity, Human Cooperation, and the Enforcement of Social Norms," *Human Nature*, Vol. 13, pp. 1–25.

Fehr, Ernst and Simon Gächter (1998) "Reciprocity and economics: The economic implications of Homo Reciprocans," *European Economic Review*, Vol. 42, pp. 845–859, DOI: http://dx.doi.org/10.1016/S0014-2921(97)00131-1.

——— (2000) "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspectives*, Vol. 14, pp. 159–181.

——— (2002) "Altruistic punishment in humans," *Nature*, Vol. 415, pp. 137–140, DOI: http://dx.doi.org/10.1038/415137a.

Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl (1998) "Gift exchange and reciprocity in competitive experimental markets," *European Economic Review*, Vol. 42, pp. 1–34, DOI: http://dx.doi.org/10.1016/S0014-2921(96)00051-7.

Fehr, Ernst and Klaus M. Schmidt (1999) "A theory of fairness, competition, and cooperation," *The Quarterly Journal of Economics*, Vol. 114, pp. 817–868, URL: http://qje.oxfordjournals.org/content/114/3/817.shorthttp://www.jstor.org/stable/2586885, DOI: http://dx.doi.org/10.1162/003355399556151.

Fischbacher, Urs (2007) “Z-Tree: Zurich toolbox for ready-made economic experiments,” *Experimental Economics*, Vol. 10, pp. 171–178, DOI: http://dx.doi.org/10.1007/s10683-006-9159-4.

Frean, M. (1996) “The evolution of degrees of cooperation,” *Journal of Theoretical Biology*, Vol. 182, pp. 549–559, URL: http://www.sciencedirect.com/science/article/pii/S0022519396901940, DOI: http://dx.doi.org/10.1006/jtbi.1996.0194.

Fredriksson, Per G and Daniel L Millimet (2007) “Legislative organization and pollution taxation,” *Public Choice*, Vol. 131, pp. 217–242, DOI: http://dx.doi.org/10.1007/s11127-006-9114-0.

Friedman, Daniel, Steffen Huck, Ryan Oprea, and Simon Weidenholzer (2015) “From imitation to collusion: Long-run learning in a low-information environment,” *Journal of Economic Theory*, Vol. 155, pp. 185–205, URL: http://linkinghub.elsevier.com/retrieve/pii/S0022053114001483, DOI: http://dx.doi.org/10.1016/j.jet.2014.10.006.

Friedman, James W. (1971) “A Non-cooperative Equilibrium for Supergames,” *The Review of Economic Studies*, Vol. 38, pp. 1–12, URL: http://restud.oxfordjournals.org/lookup/doi/10.2307/2296617, DOI: http://dx.doi.org/10.2307/2296617.

Garbarino, Ellen and Robert Slonim (2009) “The robustness of trust and reciprocity across a heterogeneous U.S. population,” *Journal of Economic Behavior and Organization*, Vol. 69, pp. 226–240, DOI: http://dx.doi.org/10.1016/j.jebo.2007.06.010.

Gintis, Herbert, Samuel Bowles, Robert Boyd, and Ernst Fehr (2003) "Explaining altruistic behavior in humans," *Evolution and Human Behavior*, Vol. 24, pp. 153–172, DOI: http://dx.doi.org/10.1016/S1090-5138(02)00157-5.

Goeree, Jacob K., Charles A Holt, and Thomas R. Palfrey (2002) "Quantal Response Equilibrium and Overbidding in Private-Value Auctions," *Journal of Economic Theory*, Vol. 104, pp. 247–272, DOI: http://dx.doi.org/10.1006/jeth.2001.2914.

Greiner, Ben (2015) "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, Vol. 1, pp. 114–125, DOI: http://dx.doi.org/10.1007/s40881-015-0004-4.

Hofbauer, Josef and William H. Sandholm (2002) "On the global convergence of stochastic fictitious play," *Econometrica*, Vol. 70, pp. 2265–2294, URL: http://discovery.ucl.ac.uk/181004/, DOI: http://dx.doi.org/10.1111/j.1468-0262.2002.00440.x.

Huck, Steffen, Johannes Leutgeb, and Ryan Oprea (2017) "Payoff information hampers the evolution of cooperation," *Nature Communications*, Vol. 8, p. 15147, URL: http://www.nature.com/doifinder/10.1038/ncomms15147, DOI: http://dx.doi.org/10.1038/ncomms15147.

Huck, Steffen, Hans-Theo Normann, and Jörg Oechssler (1999) "Learning in Cournot Oligopoly–An Experiment," *The Economic Journal*, Vol. 109, pp. C80–C95.

——— (2003) "Zero-Knowledge Cooperation in Dilemma Games," *Journal of Theoretical Biology*, Vol. 220, pp. 47–54, DOI: http://dx.doi.org/10.1006/jtbi.2003.3129.

——— (2004) "Through Trial and Error to Collusion," *International Economic Review*, Vol. 45, pp. 205–224.

Huck, Steffen and Jörg Oechssler (1999) "The Indirect Evolutionary Approach to Explaining Fair Allocations," *Games and Economic Behavior*, Vol. 28, pp. 13–24, URL: http://discovery.ucl.ac.uk/45771/, DOI: http://dx.doi.org/DOI:10.1006/game.1998.0691.

Judd, Kenneth L (1998) *Numerical Methods in Economics*, Cambridge, MA: MIT Press, pp.633.

List, John A., Azeem M. Shaikh, and Yang Xu (2016) "Multiple Hypothesis Testing in Experimental Economics," URL: http://www.nber.org/papers/w21875, DOI: http://dx.doi.org/10.3386/w21875.

McKelvey, Richard D and Thomas R Palfrey (1995) "Quantal response equilibria for normal form games," *Games and Economic Behavior*, Vol. 10, pp. 6–38, DOI: http://dx.doi.org/10.1006/game.1995.1023.

Nowak, Martin A. (2006) "Five Rules for the Evolution of Cooperation," *Science*, Vol. 314, pp. 1560–1563, DOI: http://dx.doi.org/10.1126/science.1133755.

Nowak, Martin A. and Karl Sigmund (1993) "A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game," *Nature*, Vol. 364, pp. 56–58, DOI: http://dx.doi.org/10.1038/364056a0.

——— (1998) "Evolution of indirect reciprocity by image scoring," *Nature*, Vol. 393, pp. 573–577, URL: http://dx.doi.org/10.1038/31225, DOI: http://dx.doi.org/Doi10.1038/31225.

Oechssler, Jörg (2002) "Cooperation as a result of learning with aspiration levels," *Journal of Economic Behavior and Organization*, Vol. 49, pp. 405–409, DOI: http://dx.doi.org/10.1016/S0167-2681(02)00013-6.

Offerman, Theo, Jan Potters, and Joep Sonnemans (2002) "Imitation and Belief Learning in an Oligopoly Experiment," *Review of Economic Studies*, Vol. 69, pp. 973–997.

Pettit, James, Daniel Friedman, Curtis Kephart, and Ryan Oprea (2014) "Software for continuous game experiments," *Experimental Economics*, Vol. 17, pp. 631–648, DOI: http://dx.doi.org/10.1007/s10683-013-9387-3.

Prat, Andrea and Aldo Rustichini (2003) "Games played through agents," *Econometrica*, Vol. 71, pp. 989–1026, DOI: http://dx.doi.org/10.1111/1468-0262.00437.

Rabin, Matthew (1993) "Incorporating fairness into game theory and economics," *American Economic Review*, Vol. 83, pp. 1281–1302.

Rassenti, Stephen J., Vernon L. Smith, and Robert L. Bulfin (1982) "A Combinatorial Auction Mechanism for Airport Time Slot Allocation," *The Bell Journal of Economics*, Vol. 13, pp. 402–417, DOI: http://dx.doi.org/10.2307/3003463.

Roberts, Gilbert and Thomas N. Sherratt (1998) "Development of cooperative relationships through increasing investment," *Nature*, Vol. 394, pp. 175–179.

Rockenbach, Bettina and Manfred Milinski (2006) "The efficient interaction of indirect reciprocity and costly punishment.," *Nature*, Vol. 444, pp. 718–723, DOI: http://dx.doi.org/10.1038/nature05229.

Selten, Reinhard, Klaus Abbink, and Ricarda Cox (2005) "Learning direction theory and the winner's curse," *Experimental Economics*, Vol. 8, pp. 5–20, DOI: `http://dx.doi.org/10.1007/s10683-005-1407-5`.

Trivers, Robert L. (1971) "The Evolution of Reciprocal Altruism," *The Quarterly Review of Biology*, Vol. 46, pp. 35–57.

Vega-Redondo, Fernando (1997) "The Evolution of Walrasian Behavior," *Econometrica*, Vol. 65, pp. 375–384.

Wahl, L M and Martin A Nowak (1999a) "The Continuous Prisoner's Dilemma: I. Linear Reactive Strategies," *Journal of theoretical biology*, Vol. 200, pp. 307–321, URL: `http://www.ncbi.nlm.nih.gov/pubmed/10527721`.

——— (1999b) "The continuous Prisoner's dilemma: II. Linear reactive strategies with noise.," *Journal of theoretical biology*, Vol. 200, pp. 323–38, URL: `http://www.ncbi.nlm.nih.gov/pubmed/10527721`, DOI: `http://dx.doi.org/10.1006/jtbi.1999.0997`.

Wedekind, Claus and Manfred Milinski (2000) "Cooperation Through Image Scoring in Humans," *Science*, Vol. 288, pp. 850–852, URL: `http://science.sciencemag.org/content/288/5467/850`, DOI: `http://dx.doi.org/10.1126/science.288.5467.850`.

Appendix A

A.1 Supplementary Analysis

In the treatments with 2 and 4 seconds per period similar procedures as in the treatment with 8 seconds per period were applied. Total time spent in the environment was kept constant at 80 minutes but subjects interacted with a higher frequency. The payoff function in a period was the same as before but points were converted to euro by dividing by 4000 and 8000 respectively. Subjects in both treatments exhibit behaviour similar to the behaviour observed in the original treatment with 8 seconds per period. For comparability we choose cut-offs that are equivalent in clock time to the cut-offs in the treatment with 8 seconds, i.e. periods 50 and 600 in the treatment with 4 seconds and periods 100 and 1200 in the treatment with 2 seconds. In the naive phase subjects start out with significantly higher choices in NoInfo than in Info in both treatments (two-sided MWU tests, $p < 0.001$). In the learning phase again choices start to drop massively. In the long run then subjects in NoInfo choose significantly more cooperate outcomes in both treatments (two-sided MWU tests, $p < 0.1$). Though this difference is significant in both cases, average behaviour is not as close to the one-shot Nash in either case as it is in the Info treatment reported in the main text. This is due to increased heterogeneity (note the high standard deviations reported in the captions of figures 17 and 18): though most groups converge to Nash-like behaviour, a minority of groups in each case manage to move away from Nash into collusive territory.

A.2 Supplementary Figures

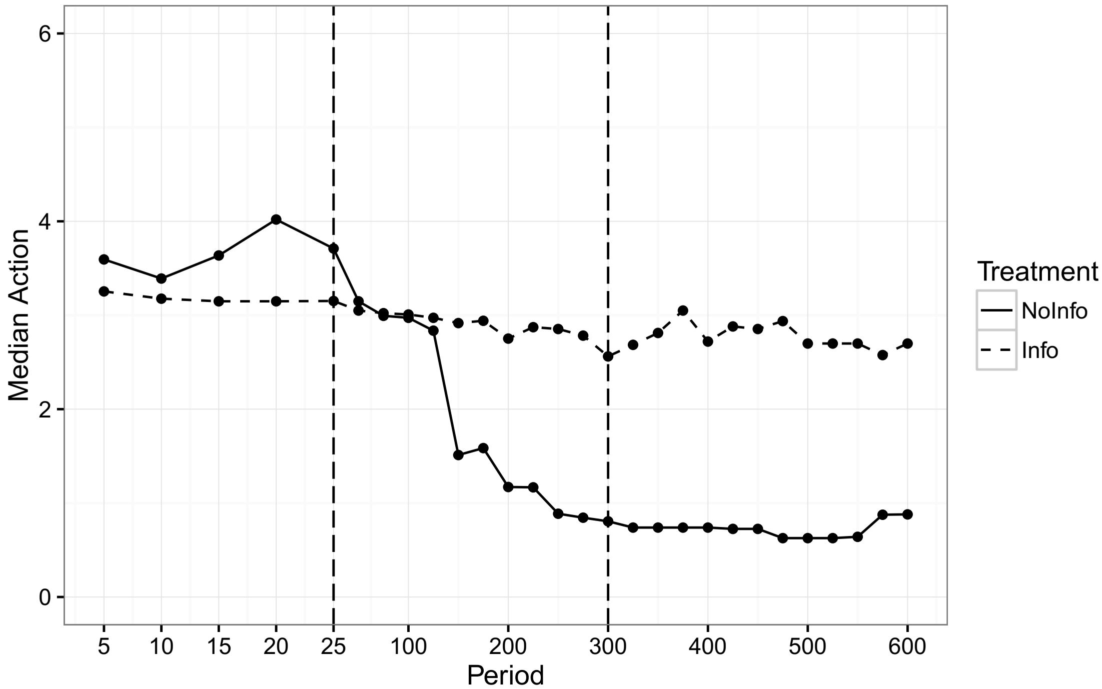


Figure 17: Median Actions with 2 Seconds per Period
For comparability the cut-offs chosen are equivalent in clock time to the cut-offs in the treatment with 8 seconds. Up to period 100 each dot represents the median action in bins of 20 periods. From period 100 (dashed vertical line) on each dot represents the median action in bins of 100 periods. In the naive phase (periods 1-100) the average action (standard deviation in parentheses) is 3.89 (1.44) in NoInfo and 3.06 (1.16) in Info. In the learning phase (periods 101-1200) this changes to 2.06 (2.04) in NoInfo and 2.42 (1.42) in Info. Finally, in the long-run phase (periods 1200-2400) behaviour settles down at 1.36 (1.71) in NoInfo and 1.75 (1.47) in Info.

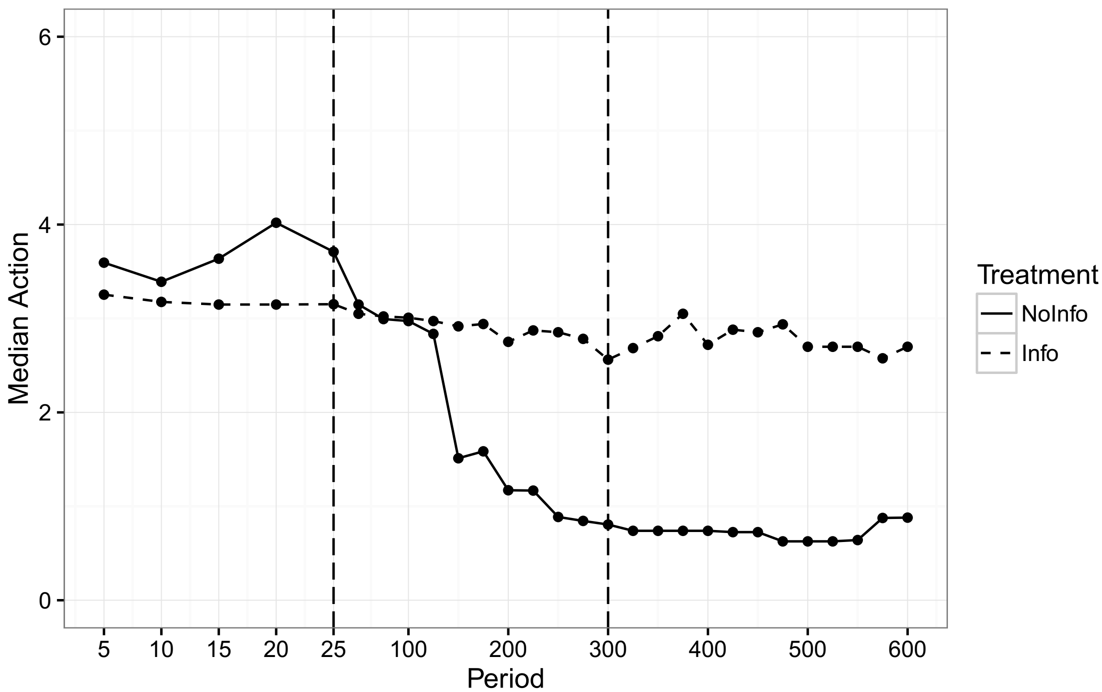


Figure 18: Median Actions with 4 Seconds per Period

For comparability the cut-offs chosen are equivalent in clock time to the cut-offs in the treatment with 8 seconds. Up to period 50 each dot represents the median action in bins of 10 periods. From period 50 (dashed vertical line) on each dot represents the median action in bins of 50 periods. In the naive phase (periods 1-50) the average action (standard deviation in parentheses) is 3.86 (1.51) in NoInfo and 3.33 (1.18) in Info. In the learning phase (periods 51-600) this changes to 2.54 (1.94) in NoInfo and 2.74 (1.71) in Info. Finally, in the long-run phase (periods 601-1200) behaviour settles down at 1.49 (1.49) in NoInfo and 2.05 (1.75) in Info.

A.3 Supplementary Tables

	Dependent Variable CHANGE		
	(1) Periods 1-25	(2) Periods 26-300	(3) Periods 300-600
COPY UP	0.422***	0.325***	0.219**
	(0.144)	(0.094)	(0.084)
COPY DOWN	0.015	-0.135	-0.320***
	(0.172)	(0.105)	(0.113)
WCLR	0.062*	0.116**	0.097
	(0.031)	(0.042)	(0.067)
BR	0.114***	0.099**	0.082
	(0.037)	(0.042)	(0.073)
INFO	0.355*	-0.152	-0.239*
	(0.186)	(0.129)	(0.136)
COPY UP:INFO	-0.518***	0.088	0.161
	(0.168)	(0.122)	(0.114)
COPY DOWN:INFO	-0.362*	0.213	0.365**
	(0.196)	(0.132)	(0.137)
WCLR:INFO	-0.006	-0.061	-0.068
	(0.053)	(0.052)	(0.075)
BR:INFO	0.226***	0.107	0.110
	(0.066)	(0.063)	(0.085)
Constant	0.264	0.278***	0.422***
	(0.161)	(0.099)	(0.121)
Observations	1,428	14,300	11,916
R-squared	0.233	0.243	0.260
F-Statistic	47.91***	138.03***	35.66***

Standard errors clustered on groups in parentheses.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 13: Linear Probability Model, Three Phases

A.4 Instructions

This section contains the instructions that subjects received in the Info treatment. The instructions are translated from German for review purposes. Instructions in the NoInfo treatment are the same except for a different screenshot and minus the sentence mentioning the black line.

Welcome! Thank you for participating in this economic experiment. If you read these instructions carefully, you can earn a non-trivial amount of money. The money that you earn during the course of this experiment will be paid to you in cash at the end of the last period. Please remain quiet and do not look at the screens of the other participants. If you have questions or if you need help, please give us a hand sign and we will come to your place. If you disrupt the experiment by speaking, laughing, etcetera, we will exclude you from the experiment without payment. We expect and appreciate your cooperation. All procedures in the experiment will take place exactly as they are described in these instructions.

Basic structure of the experiment

In this experiment the computer will match you anonymously with another player. The experiment is divided into periods. In each period you and the other player will secretly choose actions. The combination of actions that you and your partner have chosen at the end of the period will determine the amount of points that you earn in this period. We will not explain to you exactly how your points are calculated, but here are some hints: Your points in each period are determined solely by your strategy and the strategy of your counterpart. The function that determines your points will not change during the experiment. If you and your counterpart choose the same actions at some points in time A and B, you will earn the same amount at point A as in point B. Your payoff function is symmetric to your counterpart's function. If you and your counterpart choose the same action in the same period, you will earn the same amount of points.

Computer Display

Figure 19 shows the display which you will use to make choices and through which you will interact with your counterpart. At the top of the display you see a progress bar that shows how much time has passed in the current period. When the bar is full the period ends and another period starts immediately. Your action is the position (from the left to the right) of the black square at the lower part of the display. During a period you can change your preliminary action freely by moving the square like a slider to the left and to the right, or by clicking on the desired position. Your actual action for the entire period is only determined by the position of the slider at the end of the period.

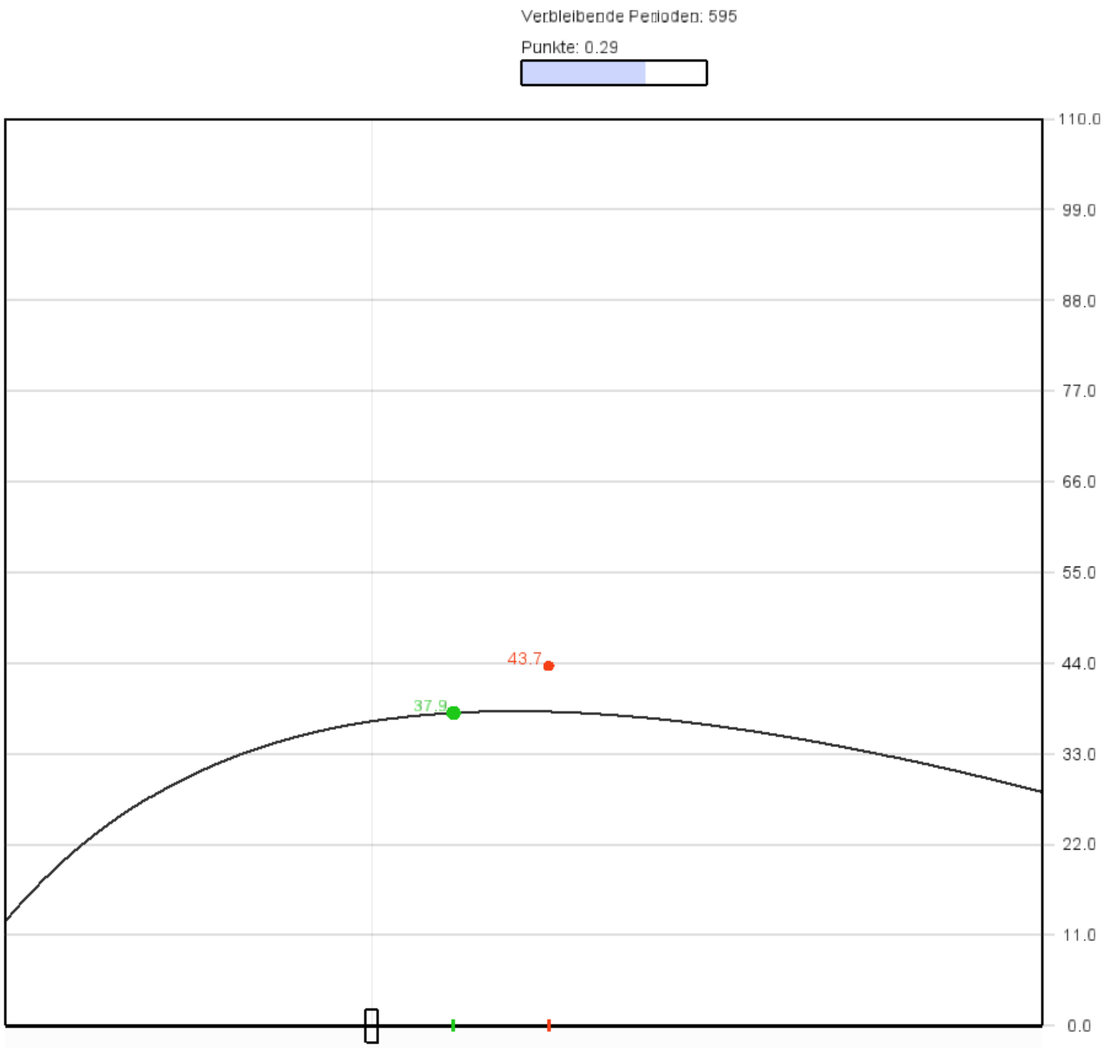


Figure 19: The Interface

After a period has ended you will see a green point that shows the amount of points that you earned in the previous period. The higher the point the more points you have earned. The exact number of points is shown next to the point. At the same time you will see a red mark at the bottom of the display which shows the action of your counterpart in the previous period. You will also see a red point that shows the amount of points that your counterpart earned in the previous period. Next to this point you will also see the number of points that your counterpart earned. Finally you will see a black line that shows you how many points you could have earned at each position of the slider, depending on the action that your counterpart has chosen in the previous period. It is important for you to understand that the action of your counterpart, your points, and the points of your counterpart are always the results of the previous period. You will not receive any information about the points or the action of your counterpart in the current period, until it has passed.

Earnings

In this experiment you will first earn points that are then converted into Euro at a rate of 0.3 Euro per point and paid out to you in cash. The exchange rate is noted down on the whiteboard at the end of the room. The earnings that are shown to you at the end of the period are the amount of points that you would earn in the entire experiment if you and your counterpart would decide the same in all periods. Your points will accumulate over the course of the experiment. The points that you have already earned are shown at the upper end of the display.

If you have not understood something, please raise your hand. We will answer your questions personally.

Thank you for your participation!

Appendix B

B.1 Supplementary Tables

VARIABLES	AGREEMENT
M/F GROUP	-1.05
	(0.56)
F/F GROUP	-1.87**
	(0.70)
MIXED MATH	0.26
	(0.68)
GOOD MATH	0.97
	(0.72)
MIXED LANG	-0.39
	(0.48)
4 SECONDS	-1.50**
	(0.58)
8 SECONDS	-2.31***
	(0.62)
INFO	-0.80
	(0.47)
Constant	2.39**
	(0.87)
AIC	133.32
BIC	157.46
Log Likelihood	-57.66
Observations	108

Standard errors in parentheses
$^{*}p < 0.02, ^{**}p < 0.01, ^{***}p < 0.002$

Table 14: Logit Regression on Finding an Agreement

	Dependent Variable CHANGE				
	8 Seconds Treatment			All Treatments	
VARIABLES	M1	M2	M3	M4	M5
COPY UP	0.274***	0.243***	0.328***	0.253***	0.254***
	(0.0769)	(0.0707)	(0.0656)	(0.0367)	(0.0368)
COPY DOWN	-0.221**	-0.144*	-0.0557	-0.00611	-0.00657
	(0.0821)	(0.0811)	(0.0650)	(0.0399)	(0.0402)
WCLR	0.108**	0.128***	0.137***	0.0801***	
	(0.0485)	(0.0407)	(0.0371)	(0.0133)	
BR	0.0871*	0.0878**	0.0696**	0.0920***	0.120***
	(0.0466)	(0.0393)	(0.0314)	(0.0164)	(0.0171)
INFO	-0.196*	-0.191*	-0.0970	-0.0881*	-0.0908*
	(0.109)	(0.101)	(0.0908)	(0.0504)	(0.0508)
COPY UP:INFO	0.124	0.140	0.0370	0.0724	0.0731
	(0.106)	(0.0999)	(0.0924)	(0.0553)	(0.0561)
COPY DOWN:INFO	0.285**	0.266**	0.165	0.0689	0.0712
	(0.109)	(0.106)	(0.0975)	(0.0582)	(0.0593)
WCLR:INFO	-0.0651	-0.0663	-0.0645	-0.00541	
	(0.0573)	(0.0499)	(0.0448)	(0.0179)	
BR:INFO	0.111*	0.102*	0.103**	0.0192	0.0161
	(0.0617)	(0.0556)	(0.0495)	(0.0290)	(0.0278)
NICE OWN (t-1)		0.0662	0.0523	-0.0503**	-0.0287
		(0.0392)	(0.0378)	(0.0213)	(0.0208)
NICE OTH (t-1)		-0.195***	-0.144***	-0.162***	-0.123***
		(0.0345)	(0.0287)	(0.0158)	(0.0178)
NICE OWN (t-2)			-0.179***	-0.223***	-0.221***
			(0.0218)	(0.0155)	(0.0155)
NICE OTH (t-2)			-0.0413*	0.000105	0.00224
			(0.0227)	(0.0150)	(0.0153)
4 SEC				0.0396**	0.0412**
				(0.0166)	(0.0163)
8 SEC				0.0469***	0.0490***
				(0.0155)	(0.0151)
NICE BOTH					0.0128
					(0.0273)
NICE BOTH:INCPROFIT					-0.113***
					(0.0219)
Constant	0.347***	0.362***	0.368***	0.432***	0.448***
	(0.0818)	(0.0773)	(0.0574)	(0.0386)	(0.0382)
Observations	26,216	26,216	23,005	113,698	113,698
R-squared	0.249	0.285	0.317	0.234	0.231

Standard errors clustered on groups in parentheses.
$*p < 0.1, **p < 0.05, ***p < 0.01$

Table 15: OLS Regression on Direction of Change, All Coefficients

	Dependent Variable: CHANGE				
	8 Seconds Treatment			All Treatments	
VARIABLES	M1	M2	M3	M4	M5
COPY UP	1.256***	1.171***	1.618***	1.205***	1.196***
	(0.321)	(0.302)	(0.319)	(0.173)	(0.172)
COPY DOWN	-1.000***	-0.636*	-0.266	-0.0168	-0.0267
	(0.355)	(0.356)	(0.335)	(0.191)	(0.191)
WCLR	0.611***	0.674***	0.766***	0.396***	
	(0.236)	(0.195)	(0.194)	(0.0653)	
BR	0.468*	0.495**	0.401**	0.458***	0.599***
	(0.245)	(0.210)	(0.178)	(0.0876)	(0.0889)
INFO	-0.771	-0.797*	-0.477	-0.416*	-0.433*
	(0.503)	(0.477)	(0.520)	(0.251)	(0.250)
COPY UP:INFO	0.537	0.595	0.190	0.331	0.339
	(0.487)	(0.472)	(0.513)	(0.273)	(0.275)
COPY DOWN:INFO	1.317**	1.232**	0.896	0.318	0.332
	(0.515)	(0.516)	(0.562)	(0.294)	(0.296)
WCLR:INFO	-0.401	-0.383	-0.405*	-0.0314	
	(0.279)	(0.242)	(0.234)	(0.0899)	
BR:INFO	0.453	0.435	0.470*	0.0964	0.0799
	(0.304)	(0.284)	(0.262)	(0.154)	(0.144)
NICE OWN (t-1)		0.348	0.273	-0.258**	-0.148
		(0.215)	(0.203)	(0.111)	(0.101)
NICE OTH (t-1)		-0.999***	-0.752***	-0.780***	-0.585***
		(0.179)	(0.150)	(0.0872)	(0.102)
NICE OWN (t-2)			-0.959***	-1.085***	-1.070***
			(0.126)	(0.0898)	(0.0896)
NICE OTH (t-2)			-0.212*	-0.00681	0.00334
			(0.129)	(0.0807)	(0.0812)
4 SEC				0.206**	0.212**
				(0.0874)	(0.0857)
8 SEC				0.243***	0.252***
				(0.0829)	(0.0811)
NICE BOTH					0.0559
					(0.123)
NICE BOTH:INCPROFIT					-0.572***
					(0.0918)
Constant	-0.832**	-0.749**	-0.661**	-0.336*	-0.256
	(0.343)	(0.331)	(0.290)	(0.180)	(0.177)
AIC	29373.37	28182.02	23775.19	128441.4	128704.1
BIC	29455.11	28280.11	23887.8	128595.6	128858.3
Log Likelihood	-14676.68	-14079.01	-11873.59	-64204.69	-64336.03
Observations	26,216	26,216	23,005	113,698	113,698

Standard errors clustered on groups in parentheses.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 16: Logit Regression on Direction of Change

	Dependent Variable CHANGE	
	NoInfo	Info
VARIABLES	M6	M7
COPY UP	0.242***	0.388***
	(0.0711)	(0.0444)
COPY DOWN	-0.0473	0.109**
	(0.0778)	(0.0490)
COPY DOWN:FEMALE	0.0659	-0.0788
	(0.0858)	(0.0691)
WCLR	0.0767***	0.0536***
	(0.0148)	(0.0137)
WCLR:FEMALE	0.0160	0.0344
	(0.0154)	(0.0222)
BR	0.0660***	0.130***
	(0.0209)	(0.0248)
NICE OWN (t-1)	-0.0559**	-0.0832**
	(0.0237)	(0.0364)
NICE OTH (t-1)	-0.235***	-0.141***
	(0.0270)	(0.0266)
NICE OWN (t-2)	-0.217***	-0.223***
	(0.0198)	(0.0313)
NICE OTH (t-2)	0.0295	-0.00504
	(0.0200)	(0.0292)
FEMALE	-0.0643	0.0943
	(0.0869)	(0.0667)
COPY UP:FEMALE	0.00491	-0.0955
	(0.0795)	(0.0678)
BR:FEMALE	0.0333	-0.0268
	(0.0278)	(0.0371)
NICE OWN (t-1):FEMALE	0.00804	0.0583*
	(0.0445)	(0.0314)
NICE OTH (t-1):FEMALE	0.0886**	-0.0110
	(0.0363)	(0.0293)
NICE OWN (t-2):FEMALE	-0.0200	0.0149
	(0.0278)	(0.0278)
NICE OTH (t-2):FEMALE	-0.0114	-0.0303
	(0.0242)	(0.0215)
4 SEC	0.0265	0.0510*
	(0.0214)	(0.0255)
8 SEC	0.0220	0.0647**
	(0.0182)	(0.0243)
Constant	0.487***	0.277***
	(0.0717)	(0.0509)
Observations	51,970	61,728
R-squared	0.251	0.226

Standard errors clustered on groups in parentheses.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 17: OLS Regression on Direction of Change with Gender Dummies, All Coefficients

	Dependent Variable CHANGE	
	NoInfo	Info
VARIABLES	M6*	M7*
COPY UP	1.187***	1.915***
	(0.331)	(0.284)
COPY DOWN	-0.213	0.572*
	(0.383)	(0.311)
WCLR	0.378***	0.272***
	(0.0786)	(0.0693)
BR	0.355***	0.655***
	(0.106)	(0.126)
BR:FEMALE	0.128	-0.142
	(0.136)	(0.188)
NICE OWN (t-1)	-0.280**	-0.449**
	(0.133)	(0.190)
NICE OTH (t-1)	-1.188***	-0.699***
	(0.185)	(0.145)
NICE OWN (t-2)	-1.120***	-1.106***
	(0.150)	(0.175)
NICE OTH (t-2)	0.169	-0.0515
	(0.130)	(0.148)
FEMALE	-0.311	0.495
	(0.412)	(0.362)
COPY UP:FEMALE	-0.0171	-0.545
	(0.371)	(0.358)
COPY DOWN:FEMALE	0.307	-0.409
	(0.421)	(0.382)
WCLR:FEMALE	0.0734	0.151
	(0.0785)	(0.102)
NICE OWN (t-1):FEMALE	0.0462	0.323**
	(0.234)	(0.160)
NICE OTH (t-1):FEMALE	0.501**	-0.0204
	(0.212)	(0.145)
NICE OWN (t-2):FEMALE	-0.00283	0.107
	(0.173)	(0.139)
NICE OTH (t-2):FEMALE	-0.0797	-0.127
	(0.145)	(0.111)
4 SEC	0.132	0.270**
	(0.110)	(0.133)
8 SEC	0.117	0.332**
	(0.0944)	(0.130)
Constant	-0.0781	-1.122***
	(0.340)	(0.305)
AIC	57720.49	70145.72
BIC	57897.65	70326.33
Log Likelihood	-28840.24	-35052.86
Observations	51,970	61,728

Standard errors clustered on groups in parentheses.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 18: OLS Regression on Direction of Change with Gender Dummies, All Coefficients

B.2 Simulations

To check if my regressions can distinguish between WCLR and reciprocity I run simulations of both processes and then run regressions on the simulated data set. In both simulations I simulate behavior in groups of two agents $I = \{1, 2\}$ for periods $t \in \{1, 2, ..., 600\}$. Agent i's action in period t is denoted by $x_{i,t}$ and her partner's action in the same period t is denoted by $x_{-i,t}$. The actions of a pair in a market $(x_{i,t}, x_{j,t})$ in the first two periods $t = 1, 2$ are randomly generated from a uniform distribution $x_{i,t}, x_{j,t} \in [0.1, 6]$ to provide a starting point. From the third period on agents follow either WCLR or a reciprocity process with some error. I assume agents to move in steps of 0.1. $\pi_{i,t}$ denotes the payoff of player i in period t. $\epsilon \sim N(0, 0.01)$ is a normally distributed error term with zero mean and variance of 0.01.

WCLR: $x_{i,t} = x_{i,t-1} + 0.1 * sign(x_{i,t-1} - x_{i,t-2}) * sign(\pi_{i,t-1} - \pi_{i,t-2}) + \epsilon_{i,t}$

When following WCLR an agent evaluates her change in action and her change in profits from two periods ago to one period ago. If profits increased, she continues to adjust her action into the same direction.

RECIPROCITY: $x_{i,t} = x_{i,t-1} + 0.1 * sign(x_{-i,t-1} - x_{-i,t-2}) + \epsilon_{i,t}$

When following reciprocity an agent evaluates her partner's change in action from two periods ago to one period ago. If her partner decreased her action she reciprocates by also choosing a lower action.

	Dependent Variable CHANGE	
	Simulated Process	
VARIABLES	WCLR	RECIPROCITY
COPY UP	0.00005	0.632***
	(0.00180)	(0.0893)
COPY DOWN		0.619***
		(0.0894)
WCLR	0.956***	-0.00384
	(0.00224)	(0.00690)
BR	0.00244	0.0184**
	(0.00864)	(0.00719)
NICE OWN (t-1)	-0.00110	-0.00950
	(0.00257)	(0.00959)
NICE OTH (t-1)	-0.00355	-0.647***
	(0.00267)	(0.00611)
NICE OWN (t-2)	0.00385	-0.00839
	(0.00282)	(0.00800)
NICE OTH (t-2)	-0.00104	0.0185***
	(0.00161)	(0.00597)
Constant	0.0211**	0.178*
	(0.00867)	(0.0890)
Observations	21,492	21,492
R-squared	0.914	0.428

Standard errors clustered on groups in parentheses.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 19: OLS Regessions on Change, Simulated Data

Table 19 reports the results of OLS regressions on the direction of change, similar to the ones in table 3. Column WCLR reports the results for the simulated data set with a WCLR process[22] and column RECIPROCITY reports the results for the simulated data with a reciprocity process. In the first column WCLR successfully picks up the WCLR process in the data, as the coefficient is highly statistically significant. NICE OTH (t-1), which should pick up the reciprocity process, is statistically insignificant. In the second column the picture is reversed. Now NICE OTH (t-1) is highly statistically significant while WCLR is statistically insignificant. NICE

[22]Note that in this column COPY DOWN is omitted in the regression due to collinearity.

OTH (t-1) successfully picks up the reciprocity process that I used in the simulations. Hence, my chosen regression model should be able to distinguish between WCLR and reciprocity in the data.

Appendix C

C.1 Symmetric QRE

A principal A chooses an action out of her strategy set $X = \{x_1, ..., x_J\}$ which consists of J pure strategies. In the symmetric treatments there is no meaningful distinction between principals A and B. For example principal B's offer of $(t_R^B(U), t_C^B(R)) = (1, 0)$ for row and column agent respectively is equivalent to principal A's offer of $(t_R^A(D), t_C^B(L)) = (0, 1)$. Furthermore, all subjects perceived the game as playing as principal A. Therefore, we only need to model principal A. Her opponent principal B's strategy is constructed by rearranging the principal A's own strategy. Principal A's strategy is a vector σ_i of length J that assigns each pure strategy a probability and adds up to 1. Given σ_i we can rearrange the pure strategies and their probabilities to construct an equivalent principal B. Given this principal B's strategy and probability vectors we can calculate the expected payoff $E\pi(x)$ for each of principal A's actions. Principals choose actions with a logit choice function and a precision parameter λ. For $\lambda = 0$ a principal randomizes uniformly between all actions but as λ grows errors become smaller and smaller until behavior approaches for $\lambda \to \infty$ a Nash equilibrium. A quantal response equilibrium for a given λ is characterized as a fixed point of such a logit response function. The probability of choosing an action $x \in X$ is as follows.

$$Prob(x) = \frac{exp(\lambda E\pi(x))}{\sum_{t \in X} exp(\lambda E\pi(t))}$$

In the theoretical model the action space for offers is continuous. A logit choice models needs discrete actions though. Therefore, we approximate the model by discretizing the action space (resembling, of course, the experimental environment). Choosing the appropriate discrete step size is not trivial. The smaller the step size the bigger the computational demand. This issue is exacerbated in our case as we have a two-dimensional action space. The number of distinct actions increases nearly quadratically in the step size.[23] Furthermore in order to calculate the expected payoffs each action has to be evaluated against every possible other action. We choose a step size of 0.5 as it computes the equilibria reasonably quickly and as it is sufficient to illustrate our point.[24] We define the set of actions as the set of actions that fulfill the budget constraint G. In PD High the budget is 6, in PD Low 5, and in COORD 3.

$$X = \{(0,0),(0,0.5),...,(0,G),(0.5,0),(0.5,0.5),...,(G,0)\}$$

For our analysis we need two algorithms. The first is the equilibrium algorithm which computes a logit equilibrium for a given λ. The second is the optimization algorithm which determines the λ that fits the data best. For a given λ we use a fixed point iteration approach (see, for example, Judd (1998)). Using some probability distribution[25] as a starting point the equilibrium algorithm computes a logit response which is used as an input for another logit response and so on. This algorithm then continues until it converges.[26]

The second algorithm computes the λ and assorted equilibrium that fit the data best. Given the data the algorithm finds an equilibrium that minimizes the negative log-likelihood using the standard R optim function.[27] Our empirical data is close to

[23]For a budget G and a step size α subjects have $\frac{1}{2}(\frac{G^2}{\alpha^2}+\frac{G}{\alpha})$ feasible strategies. This leads to $\frac{1}{4}(\frac{G^2}{\alpha^2}+\frac{G}{\alpha})^2$ combinations of strategies one has to consider in order to calculate a logit response. In treatment PD High the budget is $G = 6$. For a step size of $\alpha = 0.5$ we have 6084 comparisons. For a step size of $\alpha = 0.1$ this increases to about 3.5 million comparisons. For a step size of $\alpha = 0.01$ (the actual step size in the experiment) this increases to about 32 billion comparisons.

[24]The results do not change qualitatively for a smaller step size.

[25]We use a distribution where a subject randomizes uniformly over all actions. The results are robust to using other distributions as a starting point.

[26]A sequence never perfectly converges, so we consider a sequence as converged when the sum of the absolute probability differences between the last two iterations is smaller than 0.001.

[27]We also calculate a number of QREs for an increasing λ in order to understand how the process converges. Results are provided in figure 23.

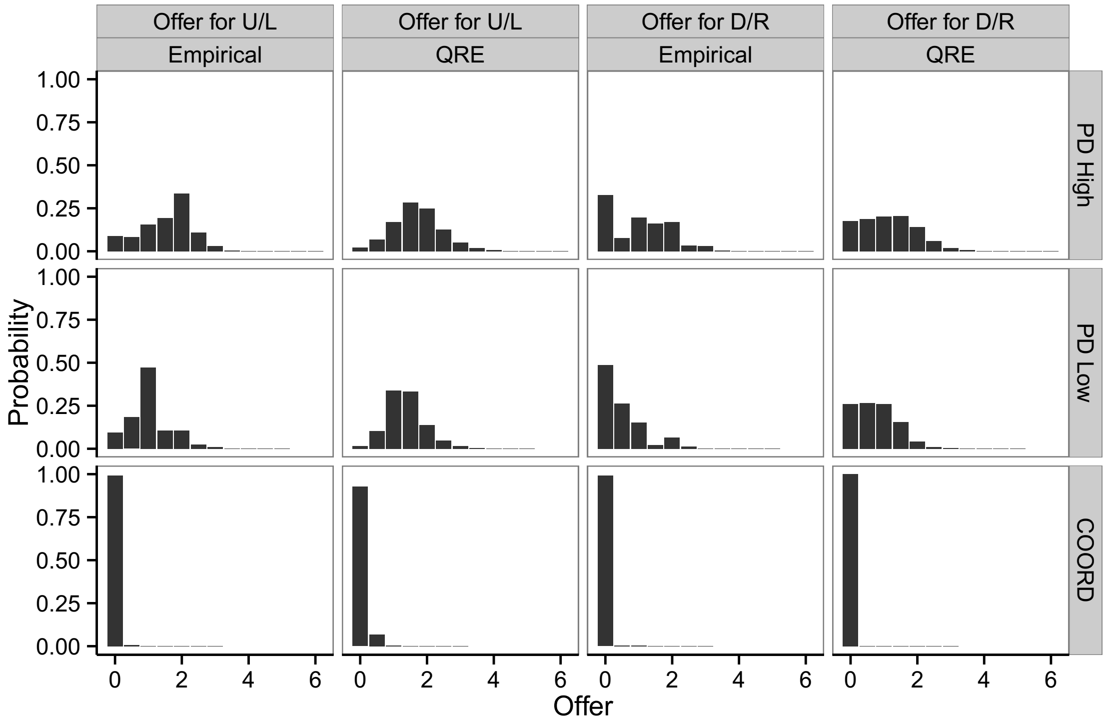


Figure 20: Histograms of Empirical and Best-Fit QRE Offers

continuity with subjects being able to specify offers in steps of 0.01. In order to fit the steps in our QRE we round each offer to its next QRE step, that is, every offer weakly smaller than 0.25 is counted as 0, every offer larger than 0.25 and weakly smaller than 0.75 is counted as 0.5 and so on. For our fits to make sense we have to assume the observations to be independent and to come from the same distribution.

C.2 Robustness Checks

The QREs are fitted on individual choice data of tuples of offers for $(U/L, D/R)$ so we check if the equilibria are also consistent with other patterns in the data. Figure 20 shows histograms for empirically observed offers and those predicted by the best-fitting QREs separately. The QRE predictions correspond nicely to the observed

		PD High		PD Low		COORD	
		Emp.	QRE	Emp.	QRE	Emp.	QRE
Offers	U/L	1.56	1.71	1.04	1.42	0.01	0.04
	D/R	1.03	1.13	0.49	0.81	0.01	0.00
Outcomes	UL	0.50	0.59	0.63	0.71	0.96	1.00
	UR	0.21	0.18	0.17	0.13	0.01	0.00
	DL	0.25	0.18	0.14	0.13	0.03	0.00
	DR	0.05	0.04	0.05	0.02	0.00	0.00

Table 20: Empirical and Best-Fit QRE Mean Offers and Outcomes

offers. Recall that we fit the data on a two-dimensional action set, whereas now we look at offers in one dimension at a time, so such a similarity does not follow by construction. The empirical mean offers are also very close to their corresponding predicted mean offers. Table 20 shows the observed and predicted mean offers. In all treatments predicted offers are reasonably close to the empirical means. The QREs also manage to capture that offers for U/L are higher than for D/R. Below the mean offers the same table also shows the observed and predicted distribution of outcomes. In all QREs outcome UL is predicted to be modal, followed by UR and DL and finally DR, just like we observe empirically. All in all, the QREs consistently fit the empirical data in multiple dimensions.

C.3 Asymmetric QRE

For the asymmetric games we calculate quantal response equilibria in a way similar to the symmetric treatments. The main difference is that due to the asymmetry in payoffs we now have to model two principals with choice probability vectors σ_i and σ_j instead of just one.[28] Also we stick to the step size of 0.5. Simple OLS regressions

[28] As before the equilibrium algorithm uses a fixed point iteration approach and consider a sequence to have converged when the sum of absolute probability difference of both vectors between the last two iterations is smaller than 0.001.

		Emp.	QRE
Offers	Principal A	1.20	1.81
	Principal B	0.90	0.92
Outcomes	UL	0.59	0.77
	UR	0.02	0.07
	DL	0.02	0.07
	DR	0.37	0.08

Table 21: Empirical and Best-Fit QRE Mean Total Offers and Outcomes in BoS

(see tables 29 and 30) show no significant time trend in any of the offers. Therefore now we consider all periods 1 to 28.[29]

As above we can now check if the equilibria are consistent with aggregate data patterns. Figures 21 and 22 again show histograms for empirically observed and best-fit QRE offers. Note that now that we consider the offers separately and thus ignore symmetry, individual offer patterns are very similar between observed and predicted offers. Tables 21 and 22 again juxtapose empirical and predicted mean offers as well as the distribution of outcomes. The predicted mean offers are quite close to the empirical means in both treatments. More importantly, principal A is predicted to offer more than principal B in either treatment, which is one feature of the empirical data. Finally, as mentioned before, in BoS the QRE cannot perfectly replicate the empirical distribution of offers because of the non-symmetric offers in the QRE. Still, the QRE predicts outcome UL to prevail most of the time, which we also observe in the data.

[29]Results do not change much if we consider periods 15 to 28 as before.

		Emp.	QRE
Offers	U	1.12	1.31
	L	1.13	1.31
	D	0.70	0.68
	R	0.65	0.68
Outcomes	UL	0.72	0.80
	UR	0.10	0.09
	DL	0.08	0.09
	DR	0.09	0.02

Table 22: Empirical and Best-Fit QRE Mean Total Offers and Outcomes in DOM

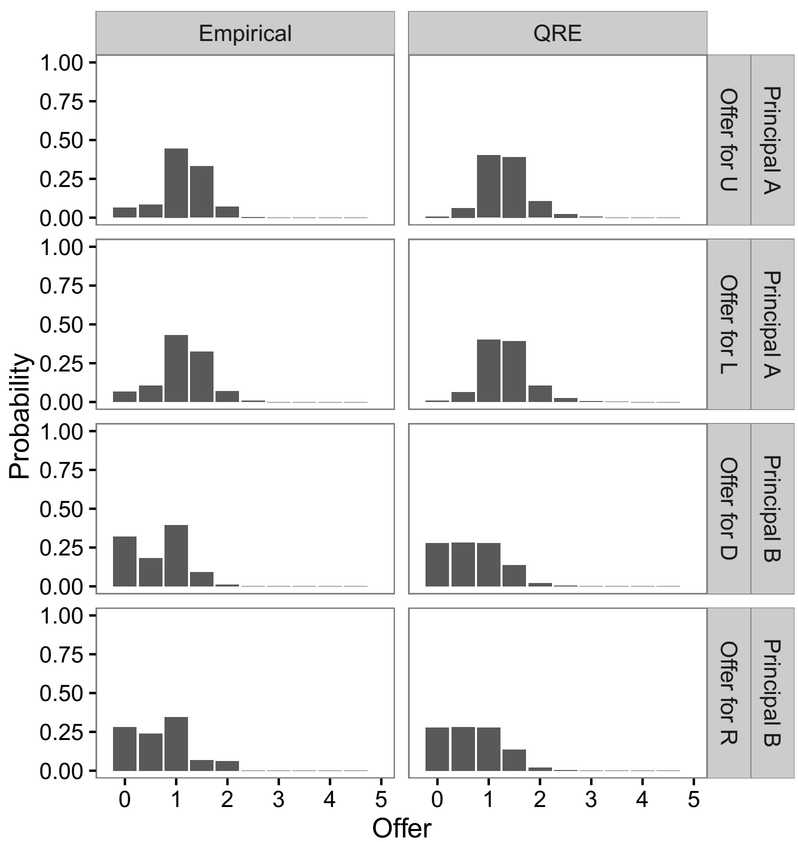


Figure 21: Histograms of Empirical and Best-Fit QRE Offers in DOM

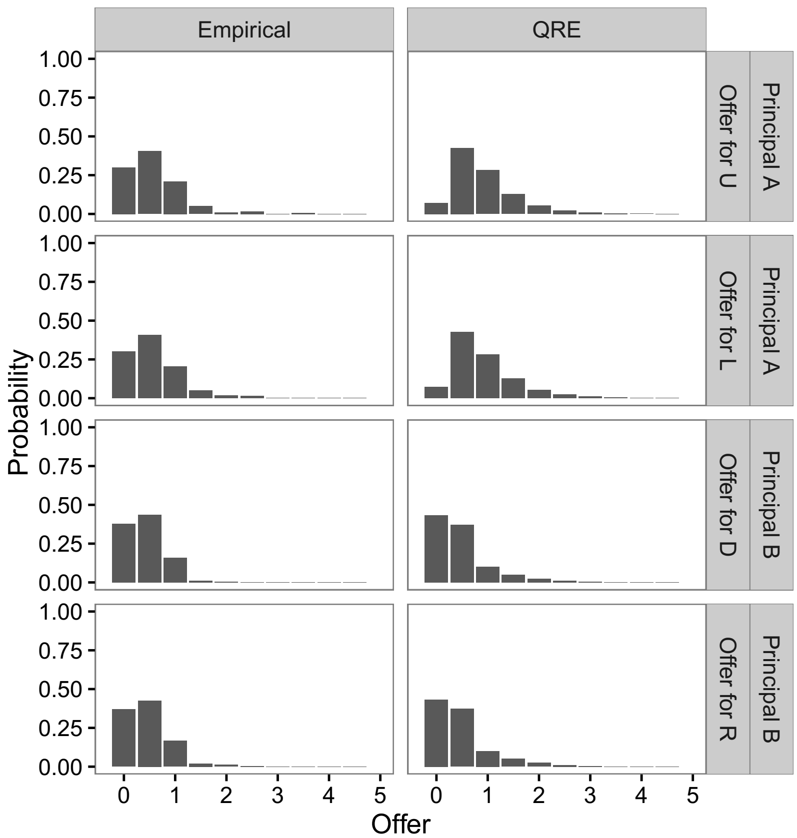


Figure 22: Histograms of Empirical and Best-Fit QRE Offers in BoS

C.4 Additional Tables and Figures

		Percent of Outcomes				Median Offer	
Treatment	Half	UL	UR	DL	DR	U/L	D/R
PD High	1^{st}	0.52	0.24	0.20	0.04	1.90	1.09
PD High	2^{nd}	0.50	0.21	0.25	0.05	1.70	1.02
PD Low	1^{st}	0.59	0.18	0.18	0.04	1.26	0.50
PD Low	2^{nd}	0.63	0.17	0.14	0.05	1.00	0.40
COORD	1^{st}	0.88	0.06	0.04	0.01	0.00	0.00
COORD	2^{nd}	0.96	0.01	0.03	0.00	0.00	0.00

Table 23: Symmetric Treatments First and Second Half

	Percent of Outcomes				Median Offer			
Half	UL	UR	DL	DR	U	L	D	R
1^{st}	0.71	0.10	0.09	0.09	1.11	1.14	0.61	0.75
2^{nd}	0.60	0.02	0.02	0.36	1.03	1.05	0.69	0.60

Table 24: DOM First and Second Half

	Percent of Outcomes				Median Total Offer	
Half	UL	UR	DL	DR	Principal A	Principal B
1^{st}	0.57	0.02	0.03	0.38	1.00	0.75
2^{nd}	0.60	0.02	0.02	0.36	1.20	1.00

Table 25: BoS First and Second Half

Change	-	NA	+	∑
+	72	5	103	180
0	216	14	75	305
-	180	0	7	187
∑	468	19	185	672

LDT

U/L PD High

Change	-	NA	+	∑
+	28	91	52	171
0	77	174	117	368
-	75	46	12	133
∑	180	311	181	672

LDT

D/R PD High

Change	-	NA	+	∑
+	74	7	69	150
0	273	11	58	342
-	169	2	9	180
∑	516	20	136	672

LDT

U/L PD Low

Change	-	NA	+	∑
+	23	58	28	109
0	68	293	112	473
-	41	44	5	90
∑	132	395	145	672

LDT

D/R PD Low

Change	-	NA	+	∑
+	3	1	1	5
0	94	544	16	654
-	13	0	0	13
∑	110	545	17	672

LDT

U/L COORD

Change	-	NA	+	∑
+	2	1	0	3
0	5	637	17	659
-	10	0	0	10
∑	17	638	17	672

LDT

D/R COORD

Note: On the x-axis LDT shows learning direction theory's predictions. The theory either prescribes a lower offer (-), a higher offer (+) or doesn't yield a prediction (NA). On the y-axis Change shows whether a subject increased (+), decreased (-) or didn't change her offer (0) in the following period. Each cell counts the number of cases. Cells are colored green (dark gray when printed in black and white) if the theory correctly predicted the change, yellow (light gray when printed in black and white) if not, and white if it didn't yield a prediction. Here, we count zeros as successes.

Table 26: Learning Direction Theory Results in Periods 14-28

	PD High		PD Low		COORD	
Offer for	U/L	D/R	U/L	D/R	U/L	D/R
Period	-0.025^{**}	-0.018	-0.038^{***}	-0.010	-0.043^{***}	-0.009^{***}
	(0.010)	(0.011)	(0.010)	(0.008)	(0.007)	(0.003)
Constant	1.920^{***}	1.246^{***}	1.649^{***}	0.715^{***}	0.505^{***}	0.129^{***}
	(0.117)	(0.133)	(0.138)	(0.107)	(0.090)	(0.046)
N	672	672	672	672	672	672
R^2	0.015	0.006	0.032	0.003	0.156	0.026

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Standard errors clustered on subject level in parentheses

Table 27: OLS Regressions With Periods 1-14

	PD High		PD Low		COORD	
Offer for	U/L	D/R	U/L	D/R	U/L	D/R
Period	0.000	-0.006	-0.005	0.005	-0.001	-0.002
	(0.010)	(0.011)	(0.006)	(0.004)	(0.001)	(0.002)
Constant	1.551^{***}	1.169^{***}	1.155^{***}	0.383^{***}	0.027	0.050
	(0.235)	(0.260)	(0.159)	(0.132)	(0.021)	(0.042)
N	672	672	672	672	672	672
R^2	0.000	0.001	0.001	0.001	0.011	0.009

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Standard errors clustered on subject level in parentheses

Table 28: OLS Regressions With Periods 15-28

	Principal A		Principal B	
Offer for	U	L	D	R
Period	-0.004	-0.001	0.001	0.001
	(0.005)	(0.003)	(0.002)	(0.002)
Constant	0.667^{***}	0.602^{***}	0.435^{***}	0.440^{***}
	(0.123)	(0.100)	(0.059)	(0.060)
N	1008	1008	1008	1008
R^2	0.003	0.000	0.000	0.000

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

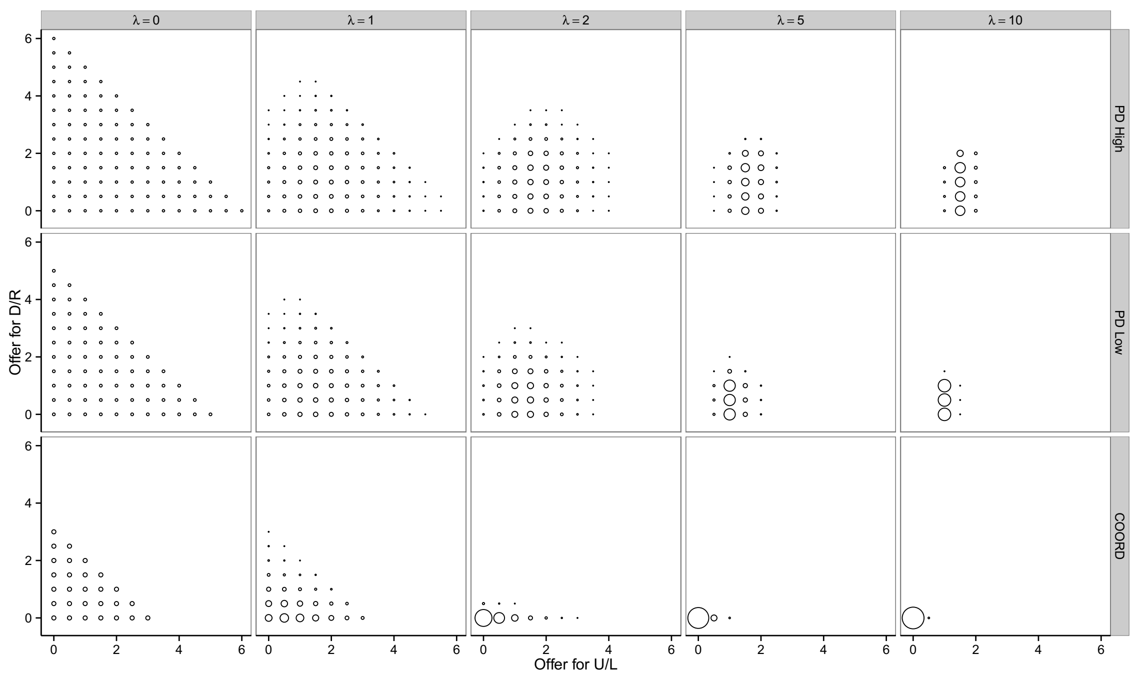
Standard errors clustered on subject level in parentheses

Table 29: OLS Regressions for BoS With Periods 1-28

	Principal A		Principal B	
Offer for	U	L	D	R
Period	0.003	-0.003	0.003	-0.003
	(0.004)	(0.003)	(0.004)	(0.004)
Constant	1.079***	1.181***	0.666***	0.695***
	(0.082)	(0.088)	(0.091)	(0.097)
N	672	672	672	672
R^2	0.002	0.004	0.001	0.003

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Standard errors clustered on subject level in parentheses

Table 30: OLS Regressions for DOM With Periods 1-28



Note: The area of a bubble corresponds to the probability of the corresponding combination of offers. Very small bubbles may not be displayed.

Figure 23: QREs of Symmetric Treatments for Increasing λ

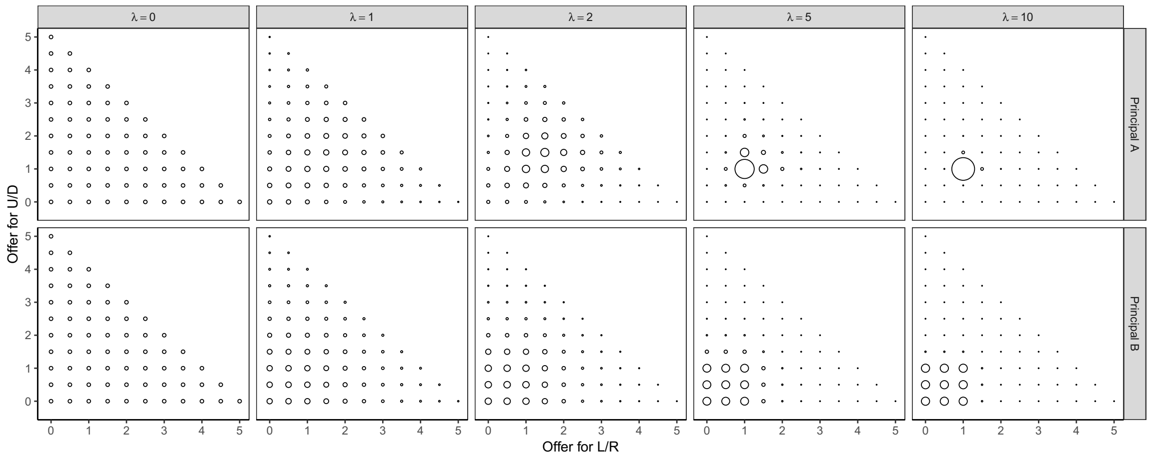
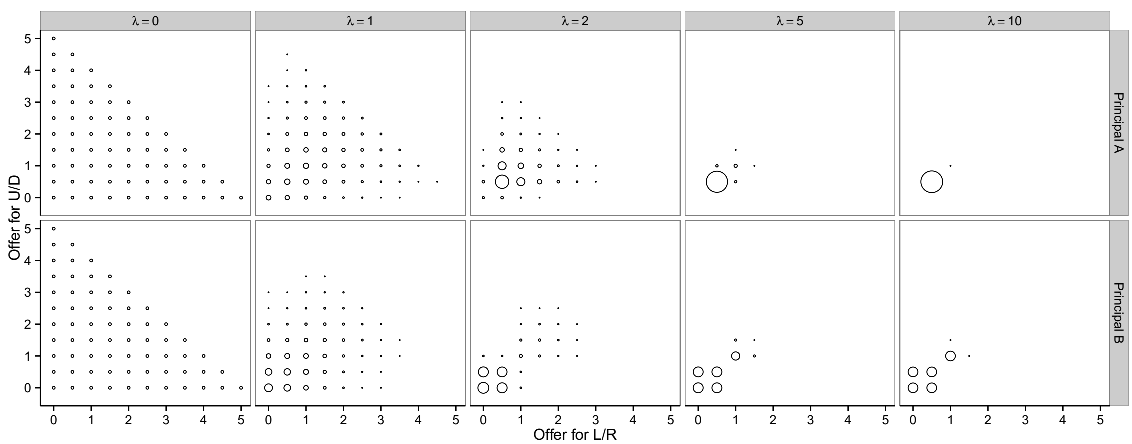


Figure 24: QREs of DOM Offers for Increasing λ



Note: The area of a bubble corresponds to the probability of the corresponding combination of offers. Very small bubbles may not be displayed.

Figure 25: QREs of BoS for Increasing λ

C.5 Instructions

These instructions are translated from the German instructions that were used in the experiment.

Welcome to our experiment!

During the experiment you may not use electronic devices or communicate with other participants. Please use only the programs and functions that are provided for the experiment. Please do not talk to other participants. If you have a question, please raise your hand. We will come to you and answer your question in private. Please do not ask your questions out loud. If the question is relevant for all participants we will repeat and answer it aloud. If you break these rules we will have to exclude you from the experiment without pay.

Rules

In this experiment you and your opponent, who is another randomly chosen participant, will decide how to pay two computer agents whose actions determine your earnings in the experiment. The computer agents both have to decide between two actions. The first computer agent has to decide between Up and Down, the second one between Left and Right. The decisions made by the computer agents determine the earnings that you and your opponent will receive. The following table shows the four possible combinations of actions by the computer agents and your winnings (the exact procedure is described in section payment). L and R stand for Left and Right, U and D for Up and Down.

	L	R
U	4;4	0;6
D	6;0	1;1

You will receive the table's first component. Your opponent will receive the second component. Therefore,

- If the first computer agent chooses U and the second L, both you and your opponent will earn 4 Euro.
- If the first computer agent chooses U and the second R, you will earn 0 Euro and your opponent will earn 6 Euro.
- If the first computer agent chooses D and the second L, you will earn 6 Euro and your opponent will earn 0 Euro.
- If the first computer agent chooses D and the second R, both you and your opponent will earn 1 Euro.

Your opponent has the exact same information.

Computer Agents

Now you will wonder how the computer agents choose their action. They choose as follows: Both you and your opponent each can promise some amount of money to the computer agents in order to choose a specific action. For this both you and your opponent have a budget of 6 Euro.

You can promise the first computer agent some amount of money to choose Down. Your opponent can promise the first computer agent some amount of money to choose Up. How does the computer agent decide? She chooses the action for which she was promised more money. If she was promised the same amount for both actions, she will choose Up.

You can promise the second computer agent some amount of money to choose Left. Your opponent can promise the second computer agent some amount of money to choose Right. How does the computer agent decide? She too chooses the action for which she was promised more money. If she was promised the same amount for both actions, she will choose Left.

Why do you offer money for D and L and your opponent for U and R? There is a reason for this. Note that it is always better for you if the first computer agent chooses D, no matter what the second computer agent does. (You will earn 6 Euro instead of 4 Euro if the second computer agent chooses L, respectively 1 Euro instead of 0 Euro if the second computer agent chooses R.) Similarly it is always better for you if the second computer agent chooses L. (You will earn 4 Euro instead of 0 Euro, respectively 6 Euro instead of 1 Euro.) In turn it is better for your opponent if the first computer agent chooses U and the second computer agent chooses R. This is why your opponent will promise money for these actions.

If a computer agent chooses an action that you promised money for, you will have to pay the amount, i.e. the promised amount is subtracted from your budget. If a computer agent chooses another auction, your opponent has to pay the amount of money she promised.

You can decide freely how much money to promise to the computer agents. The only condition is that the sum of the promised payments to both computer agents may not exceed your budget of 6 Euro. The same condition applies to your opponent.

Experimental Procedures

The situation that was described above will be repeated 30 times. Each repetition is one round. The first two rounds are trial rounds, while rounds 3 through 30 are relevant for your payouts at the end of the experiment (see section payment).

At the beginning of each round another participant will be randomly chosen by the computer and matched to you as your opponent.

At the beginning of each round you and your opponent will enter the payments that you promise to the computer agents for an action. After you and your opponent have entered the amounts the computer agents will choose their actions following the rules above. At the end of the round you will learn what payments your opponent promised and your net earnings from the round (see section payment).

Payment

Your net earnings in a round are calculated as follows. You start the round with a budget of 6 Euro. To this budget we will add the earnings that you get from the actions that the two computer agents choose (4 Euro for (U ; L), 0 Euro for (U ; R), 6 Euro for (D ; L), 1 Euro for (D ; R)). We will subtract the payments that you promised to the computer agents if they chose the corresponding action.

After the 30th and last round the computer will choose two out of the rounds 3 through 30 randomly with equal probability. Your net earnings from these rounds - and only the net earnings from these rounds - will be paid out to you in cash additionally to your participation fee.

Participation fee plus net earnings from these two rounds randomly chosen by the computer add up to the payments from these experiments, i.e. the amount of money that you will receive in cash from the experimental manager.

The money that the computer agents receive will not be paid out.

An Example

Assume that you offer 2 Euro to the first computer agent to choose D and 3 Euro to the second computer agent to choose L in the first round. Assume further that your opponent offers 2.50 Euro to the first computer agent to choose D and 2.70 to the second computer agent to choose R.

Then the first computer agent will choose U and the second computer agent will choose L. The outcome therefore is (U ; L). Therefore you net earnings in this round is your budget of 6 Euro plus earnings of 4 Euro minus the payment that you promised to the second computer agent (because she choose L). This adds up to 6 + 4 - 3 = 7

Euro. The net earnings of your opponent in this round therefore are 6 + 4 - 2.50 = 7.50 Euro.

The numbers in this example have been chosen arbitrarily. They are only intended to illustrate the rules and procedures of the experiment. They are not a suggestion to you how to decide.

Comprehension Test

Please answer all 4 questions. You may continue with the experiment only after answering all questions correctly.

Assume that you offer 0.84 to the first computer agent to choose D and 3.50 to the second computer agent to choose L. Assume further that your opponent offers 1.73 Euro to the first computer agent to play U and 2.50 to the second computer agent to choose R.

1. What will the computer agents do?

2. What are your net earnings in this round?

3. What are your opponent's net earnings?

4. What would the second computer agent have done if you had offered her 2.50 Euro as well?

(The numbers in this example have been chosen arbitrarily. They are not a suggestion to you how to decide.)